DOCUMENT RESUME

ED 347 780 FL 017 194

AUTHOR	Valdman, Albert, Ed.
TITLE	Proceedings of the Symposium on the Evaluation of Foreign Language Proficiency (Indiana University, Bloomington, March 4-6, 1987).
INSTITUTION	Indiana Univ., Bloomington. Dept. of Language Education.
PUB DATE	Mar 87
NOTE	311p.
AVAILABLE FROM	Committee for Research and Development in Language Instruction (CREDLI), Ballantine Hall 602, Indiana University, Bloomington, IN 47405 ($15.25).
PUB TYPE	Collected Works - Conference Proceedings (021)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Certification; *Communicative Competence (Languages); Evaluation Criteria; Grammar; Higher Education; High Schools; *Language Fluency; *Language Proficiency; *Language Tests; Listening Comprehension; Second Language Instruction; *Second Languages; Student Evaluation; Testing; *Test Use; Uncommonly Taught Languages; Vocabulary Development
IDENTIFIERS	*ACTFL ILR Oral Proficiency Guidelines

ABSTRACT

The proceedings of a two and one-half day symposium concerning the assessment of functional language ability through the oral interview process included the following papers: "The FSI/ILR/ACTFL Proficiency Scales and Testing Techniques: Development, Current Status, and Need Research" (John L. D. Clark, Ray T. Clifford); "The ACTFL Proficiency Guidelines: An Update" (Judith E. Liskin-Gasparro); "Problems in Examining the Validity of the ACTFL Oral Proficiency Interview" (Lyle F. Bachman); "Legal Caveats on Communicative Proficiency Testing for Graduation Requirements or Teacher Certification" (Sally Magnan); "The Extension of Language Proficiency Guidelines and Oral Proficiency Testing to the Less Commonly Taught Languages" (David Hiple); "Application of the ILR-ACTFL Test and Guidelines to the Less Commonly Taught Languages (Ron Walton); "Bennett-Biersteker Model for Proficiency Profiling in African Languages" (Sally Magnan, Patrick Bennett); "Arabic-Specific Oral Proficiency Guidelines" (Ernest N. McCarus); "The Certificate of Dutch as a Foreign Language" (Jos Nivette); "Communicative Needs: The Case of Irish" (David Singleton); "Adapting ACTFL Proficiency Guidelines to Russian" (Irene Thompson); "Application of the ILR-ACTFL Test and Guidelines to Indonesian (John U. Wolff); "Testing Listening Comprehension in the Context of the ACTFL Proficiency Guidelines (Dan Douglas); "The Problem of the Target Model in Proficiency-Oriented Foreign Language Instruction: (Albert Valdeman); "Four Challenges for Proficiency: Comments from within the AEI Proficiency Framework" (Pardee Lowe, Jr.); "Features of Pragmatic and Sociolinguistic Competence in the Oral Proficiency Interview" (Heidi Byrnes); "The Role of Grammar in the Development of Communicative Ability" (Nina Garrett); "What Price Grammar" (Simon Belasco); "L2 Vocabulary Acquisition" (Susan M. Gass); "Issues in Establishing and Maintaining a Language Proficiency Requirement (Barbara F. Freed); "Developing and Implementing Proficiency Oriented Tests for a New Language Requirement at the University of Minnesota (Dale L. Lange); and "Evaluation of Foreign Language Proficiency: Synthesis" (G. Richard Tucker). Individual presentations or group of presentations are followed by reaction/response papers, comments, and a summary of the discussion session.

ED347780

Proceedings of the Symposium on the Evaluation of Foreign Language Proficiency

March 1987
Indiana University

Edited by Albert Valdman

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

A Valdman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

FL 017194

Committee for Research and Development in Language Instruction
Ballantine Hall 602
Indiana University
Bloomington, IN 47405

BEST COPY AVAILABLE

ACKNOWLEDGEMENTS

We would like to thank all those who, directly or indirectly, have made possible these Proceedings and the Symposium which they record. John Lombardi, currently Provost at Johns Hopkins University, as Dean of the College of Arts and Sciences at Indiana University provided counsel and encouragement in the early stages of planning and concrete support in the form of a sizeable grant after the project had been formulated. Morton Lowengrub, Dean of Research and Graduate Development, obtained a guarantee of contingency internal funding while a research contract was being negotiated with the Defense Language Institute Foreign Language Center (DLIFLC), Presidio of Monterey, CA. This funding made it possible to begin planning well in advance, and thus, to secure the participation of the country's leading specialists in FL evaluation and applied linguistics early on. Ray Clifford and John Clark, in addition to helping us obtain DLIFLC funding, collaborated closely in structuring the Symposium.

Although DLIFLC provided a major part of the funding, the ideas and opinions expressed in this collective volume are those of the authors, and do not necessarily represent an official position of the DLIFLC nor any other element of the United States Department of Defense.

We would like to pay a debt of gratitude to colleagues in the various Indiana University language and area studies programs who provided significant financial support. Special thanks are due to Maurice Garnier, a sociologist in the West European Studies program, who served as de facto associate chairman of the Symposium Steering Committee. Other members of the committee were Sibylle Bolton, Harry Gradman, Eugene Eoyang and Lorraine Strasheim.

We would also like to express our appreciation to the many graduate students who insured that the Symposium proceeded smoothly by performing a myriad of tasks, including ferrying external participants from and to Indianapolis airport and serving as secretaries for the discussion sessions. Special recognition is given the staff of the Committee for Research and Development in Language Instruction, especially Helga Keller and Elena Tapia, for their indefatigable efforts and service beyond the call of duty. Elena Tapia assumed primary responsibility for copy editing and preparing the Proceedings for publication.

CONTENTS

Foreword by *Frank Medley* ix
Introduction by *Albert Valdman* xiii

The FSI/ILR/ACTFL Proficiency Scales and Testing Techniques: Development, Current Status, and Needed Research
John L.D. Clark & Ray T. Clifford 1

The ACTFL Proficiency Guidelines: An Update
Judith E. Liskin-Gasparro 19

I. The ILR-ACTFL Proficiency Test: Criterion-Reference or Achievement Test

Problems in Examining the Validity of the ACTFL Oral Proficiency Interview
Lyle F. Bachman 29

Legal Caveats on Communicative Proficiency Testing for Graduation Requirements or Teacher Certification
Sally Magnan 45

Reactions
Elana Shohamy 51

Summary of the Discussion Session
Cheryl Engber 55

II. The Application of the ILR-ACTFL Test and Guidelines to the Less Commonly Taught Languages

The Extension of Language Proficiency Guidelines and Oral Proficiency Testing to the Less Commonly Taught Languages
David Hiple 59

Application of the ILR-ACTFL Test and Guidelines to the Less Commonly Taught Languages
Ron Walton 67

Bennett-Biersteker Model for Proficiency Profiling in African Languages
Sally S. Magnan & Patrick Bennett 71

Arabic-Specific Oral Proficiency Guidelines
Ernest N. McCarus 73

The Certificate of Dutch as a Foreign Language
Jos Nivette 77

Communicative Needs: The Case of Irish
David Singleton 79

Adapting ACTFL Proficiency Guidelines to Russian: Problems, Applications and Implications
Irene Thompson 85

Application of the ILR-ACTFL Test and Guidelines to Indonesian
John U. Wolff 89

Summary of the Discussion Session
Robert Vosika & Timothy Buckwalter 93

III. Testing Listening Comprehension

Testing Listening Comprehension in the Context of the ACTFL Proficiency Guidelines
Dan Douglas 97

Response
John H.A.L. de Jong 115

Response
Harry L. Gradman 125

Summary of the Discussion Session
Cynthia A. Fox 129

IV. Sociolinguistic Aspects of Communicative Ability: The Issue of Accuracy vs. Fluency

The Problem of the Target Model in Proficiency-Oriented Foreign Language Instruction
Albert Valdman 133

Four Challenges for Proficiency: Comments From Within the AEI Proficiency Framework
Pardee Lowe, Jr. 151

Discussion of the Sociolinguistic Aspects of Communicative Ability
John U. Wolff 159

Summary of the Discussion Session
Lisa Harshbarger 163

V. Discourse Structure and Communicative Strategies in the Development of Communicative Ability

Features of Pragmatic and Sociolinguistic Competence in the Oral Proficiency Interview
Heidi Byrnes 167

Comments
Dell Hymes 179

Response
Claire Kramsch 183

Summary of the Discussion Session
Molly Wieland. 189

VI. The Role of Grammar in the Development of Communicative Ability

The Role of Grammar in the Development of Communicative Ability
Nina Garrett 193

What Price Grammar?
Simon Belasco 213

Response
Theodore V. Higgs 219

Summary of the Discussion Session
Laurie Knox 227

VII. The Role of Vocabulary in the Development of Communicative Ability

L2 Vocabulary Acquisition
Susan M. Gass 231

Discussion
James R. Child 247

Comments
Sharon Lapkin & Susanne Carroll 253

Summary of the Discussion Session
Laurel McLain 261

VIII. Practical Problems in the Administration of the ILR-ACTFL Proficiency Test and the Implementation of the Guidelines at the High School and University Level

Issues in Establishing and Maintaining a Language Proficiency Requirement
Barbara F. Freed 263

Developing and Implementing Proficiency Oriented Tests for a New Language Requirement at the University of Minnesota: Issues and Problems for Implementing the ACTFL/ETS/ILR Proficiency Guidelines
Dale L. Lange 275

Reaction
Robert C. Lafayette 291

Discussion
Sandra Savignon 295

Comments
Protase Woodford 299

Summary of the Discussion Session
Patricia J. Kyle & Laureen Hurt 303

IX. Symposium Synthesis

Evaluation of Foreign Language Proficiency: Synthesis
G. Richard Tucker 307

FOREWORD

Frank Medley
University of South Carolina

The papers selected for inclusion in this Conference Proceedings of the SYMPOSIUM ON THE EVALUATION OF FOREIGN LANGUAGE PROFICIENCY will give the reader an idea of the breadth and intensity of the discussions that formed the nucleus of the meeting in Bloomington, Indiana, in March of 1987. As a result of the prominence and expertise of the group that Dr. Valdman assembled as panelists, the ideas presented and discussed during the two and one-half days of the Symposium reflected some of the most current thinking on the assessment of functional language ability through the oral interview process. It is through discussions of this sort that the profession will make progress toward its common goal of combatting terminal monolingualism wherever it is encountered.

The involvement of ACTFL with the concept of proficiency and the subsequent development of the provisional guidelines was fueled by the President's Commission Report in 1979. As an organization dedicated to the teaching of foreign languages, ACTFL continually monitors activity at the state, national and international levels that have an impact on the discipline. Acting upon the findings of the 1979 President's Commission on Foreign Languages and International Studies, the organization began to search for ways to respond to the call of the Commission to establish a series of nationally recognized descriptors that would facilitate assessment of the proficiency of both students and teachers in existing as well as new or experimental foreign language programs. This National Criteria and Assessment Program (as it was dubbed by the Commission) should determine the language skills required to function in a variety of contexts, and should establish "language proficiency achievement goals for the end of each year of study at all levels, with special attention to speaking proficiency." But, as Hiple (1986) has observed,

> "Given the cloudiness of and the apparent conflict in the commission's charge, the profession could be expected to interpret the recommendations in various ways. The charge that had emerged seemed to serve two masters. On the one hand, it called for a document that would enable the profession to assess performance based on actual proficiency 'rather than by the number of hours spent in the classroom' (p. 13). On the other hand, this document was intended to enable the profession to set achievement goals 'for the end of each year of study at all levels.' Could a document, a common yardstick, be designed that would articulate a progressive or graduated sequence of functional language learning stages, and could the charge be met in such a way that would give special attention to speaking proficiency?" (p. 7.)

In pursuit of the elusive concept of proficiency, ACTFL obtained funding for two interrelated initiatives in 1981.

> One project, to offer oral proficiency tester training, was designed to respond to the commission's charge of assessing the proficiency of both students and teachers and was undertaken in collaboration with Educational Testing Service (ETS). The other project was designed to respond to the commission's charge of creating 'language proficiency achievement goals' and was undertaken in collaboration with the Modern Language Association.' (p. 8.)

The first ACTFL-sponsored oral proficiency tester training workshop took place in Houston in early 1982, with trainers coming from ETS and the government language schools. Thirty post-secondary teachers of French and Spanish were selected to participate from among some 300 applicants.

Simultaneous with the tester training project, ACTFL was working to produce the generic proficiency goals that would meet the challenge of the President's Commission to develop an all-purpose document that could be used both for evaluative as well as for curricular purposes. However, the form that these guidelines took was the result of several factors coming to bear on the project. First, the immediately widespread and intense interest in tester training, fueled by a strong desire on the part of teachers to have a viable way to measure oral facility "served to make the production of assessment-focused statements, particularly those referring to oral proficiency assessment, a priority." (p. 9.) As Hiple describes the situation,

> The greatest reservoir of experience in creating such assessment-focused statements resided in the government, and it was only natural that the project would draw on that expertise. The emphasis in government had been on oral assessment largely because of the living tradition of the oral interview. Since 'special attention to speaking proficiency' had been strongly recommended by the President's Commission as well, such an emphasis did not seem in violation of the charge. (p. 9)

The evolution of the guidelines was also affected by the outcomes of the Common Yardstick Project, which had endorsed the government scale, but modified so that the lower end of the scale was expanded and the upper end was collapsed into a single broad category. The result of strong interest in oral assessment, the orientation of the government personnel called on as resources, and the influence of the ETS experience all served to shape the end product.

> Thus, an assessment focused document evolved that, while addressing the four skills and culture, also served as a support document for tester training and

> oral proficiency evaluation. The achievement-focused side of the charge, the 'goals for the end of each year of study at all levels,' gradually became a secondary (if not impossible) objective of the project. In this vein, the document came to be referred to as 'guidelines,' since 'goals' was no longer an accurate description. In retrospect, then, one might say that, as ACTFL elected to defer to the President's Commission by accepting its dual concerns, an assessment-focused document became compromised by an impossible charge, namely to create 'language proficiency achievement goals for the end of each year of study at all levels.' (p. 10.)

Undoubtedly, the guidelines developed by ACTFL have both strengths and weaknesses, as does the oral interview process. Some rather glib, simplistic, and at times inaccurate statements have been made in the past four years--orally and in writing--by proponents and opponents-- regarding what both can--and cannot--do; what both are--and are not. The primary purpose of this invitational symposium was to consider the guidelines and the oral interview format and rating procedures in light of current thinking so that we can come to some accord as to where we as a profession should go from here.

Although ACTFL has been instrumental in the development of the guidelines and in the training of testers, certainly the concept of functional proficiency is neither unique to, nor the property of, the organization. On the contrary, those who have been most involved in the ongoing projects realize perhaps better than anyone else that these efforts are simply continuing the pursuit of ideas that have been with the profession for many years. The nomenclature may be different, and there are some innovative features, but the goal is still the same: find a way to fully and accurately describe--and measure-- functional ability in language. Clearly there is room for improvement. But this attempt to establish both a scale against which one's functional ability can be measured, and a process for that measurement, deserves to be given the attention not so much of the "band-wagoneers" nor of the "conscientious objectors," but of the serious scholars of language teaching and learning, whose objective is to progress from where we are to where we want to be as a profession.

It is encouraging that we are beginning to view student progress in terms other than seat time. It is encouraging that we are beginning to recognize that all skills do not develop at the same rate, but that all do contribute to the overall functional ability of the individual to use the language. And it is encouraging that we are continuing to emphasize the need to view measurement of the ability to use language as something more than knowledge of specific elements of that language. Whether these views represent dramatic innovation or logical continuation of efforts begun much earlier is not the issue. Rather the task is to determine the ultimate goal toward which we are willing to work as a profession, identify the resources at our disposal and the problems confronting us, and work together toward resolution of those problems. As the September, 1986, volume of the ADFL Bulletin suggests, in order to explore thoroughly the issue of proficiency and the profession, we must take note of the trends in the 1980s, the theoretical concerns related to the concept, and possible applications and perspectives that can be derived from an ongoing and objective consideration of the topic. It is toward this end that the

American Council on the Teaching of Foreign Languages continues to support efforts to further our understanding of the very nature of the development and evaluation of foreign language proficiency.

Reference

Hiple, D. V. (1986). A progress report on the ACTFL Proficiency Guidelines. In H. Byrnes & M. Canale (Eds.), *Defining and developing proficiency* (pp. 5-24). Lincolnwood, IL: National Textbook.

INTRODUCTION

Albert Valdman

The Symposium on the Evaluation of FL Proficiency

The ETS-ACTFL Oral Proficiency Test (OPI), derived from one in use for many years by various U.S. government agencies (ILR test), has had a wide impact on the foreign language (FL) profession over the last three or four years. There is scarcely any area of the field that has not been affected by this attempt to institute a national metric based on demonstrated proficiency in the functional use of a FL and, more importantly, to define achievement in language instruction in terms of functional use rather than exposure or command of a specific body of material.

Because it represents an attempt to modify the nature of the FL curriculum in the direction of the acquisition of functional language skills, the so-called "Proficiency Movement" should have attracted the attention of applied linguists working in the area of second language acquisition (SLA) research. Unfortunately, this has not been the case. Few of the workers in SLA research have been attracted to explore research questions surrounding the OPI, for example, the contribution of the various linguistic levels (phonology, morphosyntax, lexicon, discourse organization, etc.) to relative levels of communicative ability. On the other hand, those involved in adaptating OPI to the college and high school level have not sufficiently delineated hypotheses that could be submitted to empirical test.

The gap between development and research in the various areas surrounding the OPI and the associated guidelines is worrisome now that attempts are being made, on the one hand, to extend the notion of proficiency to curriculum design, elaboration of syllabuses, formulation of appropriate classroom procedures, and preparation of pedagogical materials and, on the other hand, to apply it to the less commonly taught languages. Without clear principles about the relationship between the criteria used to assign rankings on the OPI and the constructs underlying FL language learning, to speak of "proficiency-oriented syllabuses" or "proficiency-oriented classrooms" can only be misleading.

It is now time for constructive interaction between those involved in the Proficiency Movement and those who have been engaged in research on FL learning, particularly in the classroom context. This interaction should be mutually profitable to testing specialists and methodologists. For SLA researchers, the analysis of the verbal interactions that comprise the proficiency test represents an invaluable resource for SLA researchers, particularly those who are interested in classroom foreign language learning.

The implementation of a national metric for the assessment of functional ability in FLs poses numerous administrative and budgetary problems. Few institutions of higher learning, and even fewer school systems, are equipped to administer a resource-intensive test of functional ability. What is the most efficient administrative structure for training qualified testers and validating training of individual teachers establishing?

Clearly, the answers will vary depending on the number of learners of a given language. Whereas there are tens of thousands of learners of French, German and Spanish, only a few dozen persons each year attempt to acquire a functional ability in such languages as Dutch, Haitian Creole, or Indonesian.

Another set of questions that must be answered concerns the various uses of a national metric. At what levels of study (beginning, intermediate, advanced) should learners be administered resource-intensive tests and for what skills? Should a national metric be used for validation of foreign language study at the secondary school level? For placement in university courses? For university language requirements? For certification of language majors or prospective secondary school teachers? For admission to specialized graduate programs?

On a local level, strong support for high-level specialized discussion came from faculty associated with the language and area studies programs, most of whom represent social sciences and humanities disciplines outside linguistics and literature, who are involved in the teaching of the less commonly taught languages. Outside agencies, especially the Department of Education, are requiring these programs to implement proficiency-based evaluation in languages studied by small number of students and lacking specialists capable of responding to the challenge.

To address these various questions a Symposium on the Evaluation of FL Proficiency was held on the Bloomington Campus of Indiana University, March 4-6, 1987. As an interim step, a workshop was organized in conjunction with the 1986 annual meeting of the ACTFL in Dallas, which lay the ground for the Symposium.

Between 20 and 25 specialists were invited to prepare reports to be distributed in advance, or to serve as discussants and resource persons at the Symposium. They were selected from specialists who have in various ways been concerned with the description of communicative ability in a FL and its assessment: (1) those involved in the development and the administration of the OPI and guidelines; (2) those involved in the training of raters; (3) those involved in the elaboration of communicative-oriented curricula, syllabuses, teaching procedures, and materials; (4) those involved in the observation and analysis of second language learning. The participants represented a broad range of second and foreign language teaching situations: the commonly taught languages (French, German, and Spanish), and the less commonly taught languages (Arabic, Chinese, Japanese, Hindi, Indonesian, Russian, Subsahara African languages, indigenous languages of the Americas (including creoles), minor European languages), and the teaching of English as a second language.

In addition to the participants several overseas observers were invited, especially from Western Europe where FL teaching specialists have taken the lead in developing communicative-oriented programs and international schemes for the definition and assessment of proficiency levels, for example the Threshold Level and the Niveau Seuil. Invitations were also extended to a small number of representatives from professional groups and federal agencies involved in research and development in FL learning.

Organization of the Proceedings

These Proceedings contain the prepared papers, discussant reports, and discussion summaries of the Symposium. They consist of six parts. Part I focuses on the testing instrument developed and used by government agencies over a thirty year period and the accompanying descriptive statements relating level of performance to functional use of language (the Guidelines). In addition to providing the historical background of the Oral Proficiency Interview (OPI), the guidelines for its use as a measure of communicative ability, and the amalgam of pedagogical implications referred to as the "Proficiency Movement," this section identifies problem areas that invite empirical research and alternative forms of assessment. Part II addresses the special problems posed by assessment of communicative ability in languages not widely taught in the United States. Part III considers the assessment of listening comprehension as a competence independent of interactive communicative ability.

Part IV examines the relationship between communicative proficiency and various areas of language structure and functioning. Four areas have been identified for discussion: sociolinguistic aspects, pragmatics and discourse structure, grammar, and vocabulary. In the past, the latter two, together with phonology (which was not singled out for discussion at the Symposium), have weighed heavily in determining proficiency levels. With the impact of pragmatic and discourse oriented research on language learning the first two areas are taking an increasingly central role in determining communicative proficiency.

Part V addresses administrative and practical concerns in the adoption of a national metric for the evaluation of communicative ability on a wide scale. The two prepared papers describe pioneering attempts to set university-level FL entrance and exit requirements in terms of ACTFL-ETS proficiency levels and guidelines, whereas the discussant reports address the thorny issue of the use of these instruments for evaluating foreign language teachers' professional qualifications. All of the reports contained in Parts III, IV, and V and the Bachman paper in Section I are followed by the written versions of discussants' comments and a summary of the general discussion from the floor. The final section of these Proceedings is an agenda setting down priorities for future research and development.

Summary of Proceeding papers and discussion

PART I - The Oral Proficiency Interview and the ACTFL-ILR Guidelines – Clark and Clifford provide the historical background essential to an understanding the development of the OPI from a specific purpose evaluation instrument to a test designed to measure academic achievement. A point stressed by the authors is that in its initial form a central feature of the Foreign Service Institute (FSI) speaking test was the matching of functional and structural capabilities: the interview was carefully structured so as to probe the examinee's ability to perform the communicative tasks corresponding to a given rating level with a degree of facility and linguistic accuracy commensurate with that level. This process was later formalized by the Functional Trisection which requires a three-way simultaneous evaluation of sampled verbal performance: its linguistic accuracy, its functional breadth, and its range of topics and situations. Another key stage in the evolution of the OPI was its exportation to the academic community via the involvement of the Educational Testing Service (ETS) in a training program capable of furnishing the Peace Corps with a large corps of testers for a wide variety of languages. When it was adapted by the foreign language teaching profession, the OPI and its accompanying guidelines required major adjustments that would render it suitable for testing levels of achievement much lower than those attainable in the special purpose instructional programs devised by FSI and other government agencies. The lower end of the government scale was refined to provide a fine meshed measurement of a range of communicative ability reflecting insignificant functional capacity. For example, three distinct levels were inserted corresponding to the S-O -->S- O+ range on the FSI scale.

Clark and Clifford point out that government agencies are developing variants of the OPI that are less onerous to administer and that measure linguistic skills other than speaking ability. They also outline a program of research that would validate and provide a firmer empirical base for the OPI and the ACTFL-ILR (Interagency Language Roundtable) Guidelines. They admit that, paradoxically, the OPI falls considerably short as a face- and content-valid measure of communicative interaction, and they suggest research involving the correlation of OPI scores with a wider variety of more face- and content-valid activities and with highly specific tasks in employment and language use settings. The authors also raise the issues of intra- and inter-rater reliability in scoring OPIs at various proficiency levels and the stability of communicative ability over time.

Judith Liskin-Gasparro provides the pedagogical background for the creation of the ACTFL Provisional Guidelines and presents the reasons that led to their revision in 1986. She identifies two considerations that motivated the conflation of the upper end of the ILR scale and the expansion of the lower end. First, the CIA language school empirically validated for internal purposes a system of indicating finer shading between half-steps on the scale, for example between O+ and 1, internal. Second, in academic foreign language study assessment of high level proficiency is not practical because few students acquire proficiency above Level 3 and because, in any case, few instructors possess proficiency at the Levels 4-5 range to rate superior performance. In the revision of the ACTFL Provisional Guidelines an effort has been made to overcome a bias toward the commonly taught Western languages and a tendency to specify the use of particular grammatical structures. Instead, descriptions of levels are more generic in nature and are couched in broader functional terms. Liskin-Gasparro concludes with remarks on the "backwash" effect of the OPI and the guidelines on curricular design and teaching practices. She views these as catalysts for the adoption of communicative objectives by FL teachers and the generalization of pedagogical practices which are more likely to result in higher level of communicative language proficiency on the part of learners. The case is no doubt overstated, for the trend toward a heavier emphasis on interactive communication antedates the adoption of the OPI and the creation of the ACTFL Provisional Guidelines. It owes much more to the development of notional-functional syllabuses, themselves triggered by the extension of linguistic inquiry into interactional discourse and social aspects of language use. It is also somewhat ironic that a test designed to measure language performance independently of programs should be promoted as an agent to effect changes in pedagogical practices.

For Lyle Bachman the central question in evaluating the OPI and the ACTFL-ILR Guidelines is whether they constitute a valid measure of communicative language proficiency, defined as the ability to conduct culturally embedded verbal interaction with speakers of the target language community. He answers in the negative because he believes the OPI confounds the competence to be measured with the method of measurement. As an interactive procedure the OPI is confronted with the dilemna of trading off excessive standardization which reduces content validity but increases reliability of scoring against a larger range of sample tasks that more closely approximate realistic linguistic interaction but which reduce reliability.

Bachman views communicative language ability not as an holistic construct but as one composed of three sets of constituent capabilities, each of which should be evaluated separately, and each of which makes different predictions about an examinee's ability to perform specific linguistic tasks. The Functional Trisection, which developers of the OPI view as its keystone, becomes for Bachman its chief liability, because the very use of the same performance sample to simultaneously evaluate content, function, and structure confuses test design with evaluation, content and method. He also underscores the limited view of the nature of communicative linguistic proficiency reflected by evaluation guidelines that pay insufficient attention to pragmatic and sociolinguistic considerations. Instead, he proposes a tripartite model in which linguistic structure, narrowly defined to encompass phonology, morphosyntax, and lexicon, forms only one of the four subcomponents of organizational competence. Bachman's scheme resembles the widely used Canale-Swain model in its stress on the rhetorical, pragmatic, and sociolinguistic features of interactive verbal communication. It differs however in that it isolates as a separate component control of motor skills, grouped under physiological competence, and distinguishes more clearly between rules of language use, subsumed under organizational competence, and interactive competence. The latter identifies the ability to plan, execute, and assess socially embedded speech acts as a distinct and important part of communicative language ability.

Finally, Bachman joins Clark and Clifford in urging a broad program of research to validate the OPI and the ACTFL-ILR Guidelines. He deems it incumbent upon the developers and administrators of the test to empirically demonstrate the ability to predict specific functional performance from test scores. He insists further, that unlike the special-purpose use of the OPI by government agencies, at the school and university level, predictions must extend to domains and activities relevant to academic pursuits, including predictions about the pedagogical effectiveness of teachers.

Magnan explores some potential legal problems raised by the use of the OPI for purposes of establishing degree requirements and, particularly, for teacher certification. Referring to court cases, she stresses the need to establish the content and instructional validity of the test and its efficacy as a predictor of teaching competence. The latter link will prove difficult to establish in view of the multiplicity of objectives of general academic FL instruction and the vast array of knowledge and skills it demands on the part of teachers. For these reasons, Magnan recommends that the use of the OPI or similar instruments for the latter purpose be in the hands of professional teacher organizations.

Elana Shohamy considers the OPI a potential dangerous weapon unless serious questions of validity and reliability are resolved. She also mentions the need for applying other citeria than validity and reliability, for example, utility, accuracy, feasibility, and fairness. The latter criterion would reduce the risk of litigation. In her opinion the OPI's validity is reduced by the narrowness of the range of oral performance elicited. In contrast she cites the Israeli test of English speaking proficiency which comprises four different oral tasks ranging from an interview to a group discussion.

PART II - The application of the OPI and the Guidelines to the less commonly taught languages – This part of the Symposium took the form of panel discussion stimulated by a general report by David Hiple. In this introductory document Hiple offers statistics that show that there is a clear line of cleavage between the commonly taught FLs (CTFL): Spanish, French, and German, and other languages taught in US high schools and universities, but that the latter do not form a unified set. If enrollments in Russian total less than 25% of those in German, the former are 60 times higher than those in Swahili, the more widely studied of what Hiple labels the "much less commonly taught languages." His report describes the special problems encountered in the application of the OPI and Guidelines to both categories of less commonly taught languages (LCTFLs). These include: a small number of students; a limited pool of trained evaluators; great linguistic distance from English and, especially, a wide cultural gulf between them and mainstream middle class American culture; and finally, special sociolinguistic features, such as coexistence with a more prestigious and politically dominant language in a diglossic situation.

The panel participant reports provide insights and comments from the perspective of the teaching of a variety LCTFLs (and "much" less commonly taught) languages. The moderator, Ronald Walton, opened the discussion by identifying some of the key differences between the evaluation of LCFLs and that of the CTFLs. In addition to those mentioned above they include differences in teaching personnel--in particular the disproportionate number of native speakers--and a much longer training period to attain comparable levels of proficiency. He expressed his unease with the guidelines for the levels at the lower end of the scale, a recurrent concern at the Symposium. Sally Magnan describes the Bennett-Biersteker multi-componential model devised for African languages. That testing model assigns heavy weight to sociocultural factors in the evaluation of communicative language ability, and it has a major diagnostic focus. After discussing the special problems posed by the need for learners of Arabic to demonstrate proficiency in varieties of the language that differ significantly at all linguistic levels, Ernest McCarus underscores the diagnostic function of proficiency testing and the need to provide feedback to learners. David Singleton brings to the discussion both the Council of Europe communicative approach perspective, which defines curriculum in terms of learner needs, and consideration of the special problems of teaching moribund languages, such as Irish in Ireland, which are no longer media of communication in the strict sense. He argues for a broader definition of communication so that it covers the case of languages learned for access to a certain cultural heritage and for reception as well as production. Jos Nivette describes a proficiency test for Dutch developed jointly by the governments of the Netherlands and Belgium for use mostly in the latter country. In reviewing the development of guidelines for Russian, Irene Thompson introduces the novel view that they represent an attempt to capture salient features of interlinguistic development at various crucial stages. John Wolff addresses the issue of feasibility in the adaptation of the guidelines to the teaching of Indonesian. In view of the small number of students and the paucity of potential competent raters he advocates the administration of a prerecorded variant of the OPI similar to procedures devised at the Defense Language Institute or a speaking proficiency test being developed in conjunction with the use of a proficiency-based test to set entrance standards at the University of Minnesota.

PART III - Testing listening comprehension – Dan Douglas opens his article with an inventory of listening comprehension tests currently available. On the basis of task type involved he distinguishes five categories: simple listening and answering questions on the text (usually multiple-choice or true-false); reduced redundancy; repetition and imitation, in particular dictation; simulation of conversational interaction; and media transfer. In a brief critique of the 1986 ACTFL Guidelines for listening comprehension, he singles out the need to validate the underlying constructs for each level on the scale across the three components of the Functional Trisection. For example, he argues that it must be shown that an examinee who "can understand the main ideas and facts from interactive exchanges" of a text "can also recognize learned materials and isolated words and phrases when strongly supported by context". In the central part of his article Douglas discusses what he perceives as the main issues in the elaboration of listening comprehension tests: criterion-referenced testing, the notion of context in listening comprehension, specific-purpose versus general-purpose testing, and the use of technology. He points out that the major problem in devising criterion-referenced language proficiency tests is, in the light of the complexity of linguistic communication, being able to assert with confidence that mastery of a set of linguistic features corresponds to a certain degree of communicative effectiveness. After reviewing current thinking on the notion of context in language testing, he concludes that there is need for a clearer distinction between internal and external context and for a better understanding of the relationship beween input and knowledge. As a general-purpose test, Douglas suggests, the OPI is problematic because it does not clearly define what constitutes appropriate domains of language use for classroom FL learners. For what types of listening tasks should classroom instruction prepare students, and what type of related skills and knowledge should tests assess? Finally, he finds computer assisted testing and interactive audio/video promising for the development of tests that capture the global nature of culturally embedded communicative interaction. His one caveat is that we resist the temptation of reductionism and not view language development as linear or accept the restrictions of unidimensional response theory.

In the discussion of Douglas' paper, John de Jong brings to bear his experience with current European testing programs, many of which have been profoundly influenced by the Council of Europe initiated Threshold Level movement and the Graded Objectives movement in the United Kingdom. One of his perspicacious observations requires mention. He suggests that at lower levels of proficiency learners are more dependent on structural features and that, consequently, at these levels tests might need to be criterion referenced and discrete point. More advanced learners have acquired sufficient strategic competence to compensate for structural deficiencies, and their proficiency can be assessed on a single dimension. These facts would lead us to question whether an instrument devised by FSI to evaluate a relatively high level of proficiency (S-3 on the FSI scale and Superior on the ACTFL scale) should be the model for evaluating the much lower levels of competence realistically attainable by classroom learners.

PART IV- The relationship between communicative language ability and aspects of language structure and language functioning sociolinguistic aspect – Albert Valdman's article deals with the construct of educated native speaker (ENS) speech, which constitutes the implicit target norm in the ACTFL Guidelines and the explicit one for the ILR version. He argues that in classroom FL instruction such a target is too restrictive, and he provides as an alternative a multitarget model of language variation according to which speakers differ with respect to norms and the degree of influence exercised on them by shared norms in particular social contexts. Also, native speakers' speech varies with regard to relative planning in response to another set of social and contextual factors. In addition to the well documented variation determined by the nature of linguistic tasks, FL learners' interlanguage is influenced by variable orientation toward the competing norms of the target language community. Valdman suggests that the special social context represented by the FL classroom warrants the establishment of special pedagogical norms. He sets down sociolinguistically and psycholinguistically based principles for the elaboration of these norms and illustrates their application to French WH-questions, a notoriously variable feature of that language. Because it involves the presentation in pedagogical materials of sociolinguistically stigmatized features, the notion of pedagogical norm raises the issue of fossilization. Valdman points out that sociolinguistic deviance is unavoidable in early-level interlanguage; even speech targeted on ENSs may be deviant in certain social contexts. He concludes by suggesting that fossilization may be reduced by a better match between communicative demands and linguistic means available to learners: the former are not lessened but rather the latter are fitted to the learners' limited processing capabilities.

John Wolff's comments show that cultural considerations powerfully determine the degree of sociolinguistic deviance tolerable in learners' interlanguage. Unlike the Western European languages in general, those of the Far East require speakers to constantly situate themselves socially and choose appropriate linguistic variants with respect to their interlocutors. It is clear that for such languages suitable guidelines must incorporate statements about mastering rules of language use at the lowest proficiency levels

Pardee Lowe addresses four challenges to the AEI (ACTFL-ETS-ILR) notion of proficiency which he perceives in Valdman's paper. First, that more precise definitions need to be given for such terms as "proficiency," "accuracy," and "fluency". Second, that it must be shown how the AEI framework encompasses language varieties. In this regard, Lowe believes that the notion of the ENS is a useful one in the assessment of high proficiency levels, and he draws a distinction between ENSs, nonnatives with ENS competence, and native non-ENSs. He also maintains the concept of fossilized street learners of a FL who are characterized by an extensive vocabulary but unremediable gaps in the mastery of grammatical structures. The third challenge calls for a fuller specification of the AEI developmental hierarchy. This is needed to validate the points in the OPI scale posited for beginning and intermediate high school and university FL learners. Finally, Lowe agrees that the developmental paths suggested by the notion of pedagogical norm invite exploration.

Discourse and pragmatics – Heidi Byrnes attempts to demonstrate that at a deeper level the OPI rating scales do assess examinees' mastery of discourse organization and their ability to engage successfully in a range of speech events. Although they are written in terms of structural features and appear to focus on accuracy, the Guidelines provide information about learners' development along two intersecting hierarchies related to pragmatics and discourse organization. One hierarchy ranks speech events in terms of their relative complexity, for example from simple naming to more complex negotiating. The other hierarchy orders varying means of carrying out the same speech event. This latter hierarchy comprises several axes in addition to that of relative structural complexification, for example, the degree of independence from context, the degree of creativity and initiative on the part of the learner, and the degree of register and stylistic manoeuver. Byrnes claims that at higher levels of proficiency these two hierarchies guide raters in making judgments about discourse and pragmatic appropriateness and acceptability rather than about structural accuracy.

She also provides an interesting definition of the criterion of fluency by relating it to unplanned speech in which the false starts and repetitions that result from on-line editing manifest a higher level of proficiency than seemingly fluent speech. As she puts it, "'fluent' speech is not so much flawless as it is disfluent in ways that are acceptable to the [native] listener."

Dell Hymes, the scholar with whom the notion of communicative competence is most intimately associated, stresses the holistic nature of communication and the difficulties inherent in assigning its constituent features to particular linguistic levels such as syntax, semantics, pragmatics, or sociolinguistics. He warns that the narrowness of the empirical base introduces dangerous ethnocentric bias in current discussions dealing with fields of linguistic inquiry beyond phonology and syntax. In inviting us to explore the diversity of ways of speaking, he identifies the concept of lines in narratives as a promising starting point. Hymes also rejects the distinction between native and nonnative performance that underlies the entire AEI effort in favor of a distinction between educated and a noneducated ways of speaking, without of course implying any value judgment in distinguishing them.

Claire Kramsch supports Byrne's attempt to incorporate strategic competence and sociocultural concerns in the OPI scale, but she perceives basic flaws in the instrument that make all attempts to broaden the scope of rater judgments subject to reductionism. Because of the interactional nature of sociolinguistic competence, the rater, as a cooperative partner in the elicitation of ratable samples, introduces an uncontrollable variable. As Bachman pointed out this results in a serious confounding of test content and method. Also, basing the Guidelines on specific norms is incompatible with the evaluation of sociolinguistic competence which requires constant shifting of norms in reaction to subtle changes in the social situation, changes that inevitably occur during interactional exchanges. Finally, Kramsch points out the unadmitted sociocultural bias inherent in a test that places a high premium on conversational management skills associated with certain cultural levels and social groups.

The role of grammar in communicative language ability – Grammar--in its numerous senses--has always played a central role in FL teaching. Nina Garrett tackles the difficult task of relating grammatical competence to the other constituent competences of communicative language ability. With regard to the place of grammar in the OPI and the ACTFL Guidelines, she delineates two mutually exclusive positions: (1) relative levels of grammatical accuracy and sequences for the use of specific grammatical features are determined on the basis of experiential evidence, in which case no independent validation (for example, on the basis of psycholinguistic evidence) is necessary; or (2) such independent validation is used to legitimize claims about the developmental order of grammatical features and specification of levels of accuracy. Only in the latter case may the Guidelines serve to inform syllabus design and pedagogical practice. To account for the contribution of grammar to communicative ability Garrett develops a psycholinguistic (processing) model according to which meaning, viewed comprehensively to include pragmatic, sociolinguistic, and sociocultural aspects, is mapped onto lexical and morphosyntactic features. Utterances which are generated by this mapping are compared against a set of templates representing native speaker grammatical, pragmatic, etc. competences. She stresses the fact that the processing rules this model yields are not isomorphous with those determined by linguistic analysis. They are in addition highly idiosyncratic. Garrett concludes the article with recommendations for research not tied to theories of first-language and naturalistic second- or foreign language learning but focused instead on the psycholinguistic processing specific to classroom FL instruction.

Simon Belasco comments that the problematic place of grammar in FL instruction stems from the erroneous emphasis on production rather than comprehension of FL texts. Arguing from the perspective of Chomsky's Universal Grammar he disagrees with Garrett's downgrading of first-language acquisition principles for classroom FL learning. He rejects reference to a contrived inductive base and explicit grammatical syllabuses in favor of an approach in which learners are introduced to grammatical principles of the target language as these occur in authentic samples, and as they pose problems for comprehension. These principles, which determine deep structure relations such as predicate arguments, differ markedly from traditional surface features.

Theodore Higgs starts from the assumption that the teaching of grammar is an integral part of FL instruction, and considers the only issue worthy of discussion, namely, the definition to be given to the term "grammar" in the specific FL classroom learning context. He discards the widely notion of grammar as a set of rules and formulae for the construction in favor of grammar as a system that relates meaning and forms. Pedagogically-oriented grammar should focus on more direct and transparent links between form and meaning so that students may be led to discover and use grammatical features that enable them to express express meanings needed for the communicative acts in which they are expected to engage. To what extent the target language grammatical machinery can be pruned is an empirical question in which one needs to consider both the communicative import of features and their ease of acquisition. Concerning the role of the ACTFL Guideline in guiding the acquisition of grammar, viewed from this perspectived, Higgs affirms that they "are explicitly pedagogy independent and ... inform neither methodology nor syllabus decisions."

The role of vocabulary in communicative language ability – In the AEI scale vocabulary range is the major factor in determining proficiency assessment at the lower end. Yet, as Susan Gass points out, the amount of research dealing with the acquisition of vocabulary in second language learning pales before that devoted to grammar. The lexicon has always been viewed as relatively unstructured, and research has focused on questions of input characteristics, witness the attention given to such external features as frequency and utility in determining basic pedagogically oriented vocabularies, instead of the process of acquisition. The latter orientation is central to the establishment of developmental scales and the determination of the lexicon's role in overall communicative language ability. Gass cites native speaker reaction studies that reveal the native speaker's awareness of the centrality of the lexicon in language use. Her review of research also shows that the lexicon forms a tightly structured network whose elucidation should contribute to a better assessment of learners' knowledge and control the target language vocabulary.

In his comments James Child looks at potential learning and communicative problems by applying to vocabulary a classificatory scheme identifying four ascending levels: phono-morphological, morpho-lexical, lexico-semantic, and semo-pragmatic. He points out that traditional pedagogical practice has focused on the second level since form-meaning relationships are readily amenable to treatment in the form of lists. But linguistic nucleation, which involves the integration of individual items in semantic networks, cannot be attained unless attention is paid to higher level lexical relationships. As Kramsch suggests, semo-pragmatic relations require a level of sociolinguistic competence that may defy assessment by means of the OPI.

Sharon Lapkin and Susanne Carroll start from qualitative research that suggests that younger (Grade 6) Canadian immersion learners of French show a pattern of lexical acquisition that differs qualitatively from that of native speaker counterparts. Specifically they use fewer morphologically and syntactically complex verbs and tend to use verbs with wider semantic coverage. They then outline and exemplify pedagogical procedures designed to increase the learners' capability of forming more complex form and meaning relationships.

PART V - Administrative and practical problems – This section centers around two progress reports of the attempt to institute proficiency language requirements at the university level. Barbara Freed describes the University of Pennsylvania's pioneering in 1981 of a variant of the FSI test and ILR guidelines as a proficiency-based exit requirement in fourth-semester FL courses. The initiative required a two-year preparatory period during which administrative procedures were devised, student fears allayed, suitable instruments implemented, and appropriate cut-off levels determined. Although the introduction of proficiency-based exit requirements has not led to a demonstrable improvement in student performance after two years of study, it has invigorated the FL programs and markedly improved articulation between lower-division courses by providing long-term objectives independent of specific course content. Most importantly, it has shown that an instrument such as the OPI that makes special demands for staff time and training and that induces considerable student anxiety can indeed be integrated into the curriculum of mid-size university programs.

As was the case with the University of Pennsylvania program, the University of Minnesota initiative described by Dale Lange was motivated by an educational policy decision to define achievement in FL study in terms of demonstrable proficiency rather than time spent in study. The Minnesota program is a bolder experiment because it involves a redefinition of both entrance and graduation requirements in proficiency terms. A typical midwestern state university, Minnesota has larger lower-division FL programs than those of private institutions like the University of Pennsylvania. Also, changes in the entrance requirements of public universities necessarily have a backwash effect on local high school graduation requirements, and they require close coordination between administrators and teachers at both levels. These differences are reflected in the form of the test itself and also in the guidelines that underlie the proficiency ratings and the levels of minimal performance. As was the case with Freed, Lange does not adopt the AEI Guidelines unquestioningly. On the contrary, he underscores some of their shortcomings, particularly for the reading requirement, and he recommends continued validation and revision. He concludes with a thorough and candid analysis of the problems that the implementation of any integrated proficiency-based FL program faces given the many inadequacies of prevalent teacher training, curriculum and syllabus design, material development, and, most importantly, administrative practices in major university FL programs.

Robert Lafayette opens his comments with the startling information that in 1984 the administration of the OPI to Texas FL teachers showed than half failed to score at the Advanced Level (S-2). He reports that the imposition of a communicative proficiency requirement for state certification has led to training programs, including mandatory study abroad, that promise to create a pool of teachers who meet a standard of minimal, albeit low, performance.

Sandra Savignon, although she lauds efforts such as Penn's and Minnesota's to adapt the ACTFL Guidelines to the task of devising more rigorous and accountable FL programs, expresses deep reservations about the OPI and the Guidelines. She agrees with several of the authors of these Proceedings in asking whether performance objectives implemented for special-purpose adult learners are suitable for youngsters and young adults learning a FL for general academic purposes. She also doubts that the Guidelines are truly criterion-referenced and provide an evaluation that is independent from the learning context.

Protase Woodford emphasizes the need for compromise in the generalization of a common metric for evaluating classroom learners' ability to use a FL for various functional purposes. He points out that the University of Pennsylvania's pioneering initiative required setting standards below those needed for students to reach the functional objectives selected. For instance, attainment of the Advanced Level in reading would scarcely be sufficient for the analysis of a piece of literature. He suggests the elaboration of two sets of scales: one set would provide teachers with detailed information for pedagogical purposes; a second set, couched in more general performance objectives, would be addressed to users.

PART VI - Concluding remarks – G. Richard Tucker opens his synthesis of the Symposium by remarking that much of the controversy surrounding the adaptation of the ILR guidelines to the general FL education programs of high schools and universities stems from the attempt of their framers to meet a variety of incompatible objectives and to please several types of clientele. He notes the lack of precision and various meanings of many key terms that appear in the papers and discussion, and he urges that these terms be clarified and that their definitions be based on operational criteria. For the validation of the OPI and the Guidelines he offers as a model the proactive research program that has accompanied the development and refinement of the TOEFL test, and he enumerates topics for a vigorous research agenda. Finally, Tucker stressed the urgency of disseminating information and the need for American specialists to be aware of related developments in other countries.

THE FSI/ILR/ACTFL PROFICIENCY SCALES AND TESTING TECHNIQUES: DEVELOPMENT, CURRENT STATUS, AND NEEDED RESEARCH

John L. D. Clark and Ray T. Clifford
Defense Language Institute

The major purposes of this paper are to briefly describe: the rationale and history of development of the "proficiency-based" testing movement in the United States, its current status, and areas in which additional research and development work is recommended to enhance both the psychometric validity and practical value of this approach.

1. History

"Proficiency-based" testing, as the term is used here, may be considered to have had its beginnings in the early 1950s, with a series of measurement initiatives undertaken by the Foreign Service Institute (FSI) of the U.S. Department of State, the federal agency responsible for training diplomats and other foreign service officers in foreign language and area studies relevant to their overseas posting. (See Wilds, 1975; Sollenberger, 1978.) The FSI for some time had been dissatisfied with its own and other available language tests and testing procedures, based on multiple-choice or other discrete-item approaches which, although capable of discriminating among examinees as to their relative ability to handle the particular questions included in the test, were not able to indicate in any meaningful way how well examinees scoring at a particular point on the scale would be able to carry out the particular language-use tasks they were expected to encounter in the actual assignment abroad.

Development of proficiency scales – The "breakthrough" approach adopted by senior FSI testing staff to address this problem was to develop a series of verbally-defined levels of general proficiency in speaking, these descriptions being based on the experience of FSI Washington staff and in-field service officers concerning typical language use requirements and degrees of language performance demonstrated by FSI graduates in the field. The resulting scale consisted of five paragraph-length descriptions of general speaking performance, ranging from a basic "survival" level of competence (Level 1) up to the full extent of ability demonstrated by an educated native speaker of the language (Level 5). The verbal definitions corresponding to each of the five original FSI proficiency levels are shown below:

Level 1 - ELEMENTARY PROFICIENCY. Able to satisfy routine travel needs and minimum courtesy requirements. Can ask and answer questions on topics very familiar to him; within the scope of his very limited language experience can understand simple questions and statements, allowing for slowed speech, repetition or paraphrase; speaking vocabulary inadequate to express anything but the most elementary needs; errors in pronunciation and grammar are frequent, but can be understood by a native spaker used to dealing with foreigners attempting to speak his language; while topics which are "very familiar" and elementary needs vary considerably from individual to individual, any person at Level 1 should be able to order a simple meal, ask for shelter or lodging, ask and give simple directions, make purchases, and tell time.

Level 2 - LIMITED WORKING PROFICIENCY. Able to satisfy routine social demands and limited work requirements. Can handle with confidence but not with facility most social situations including introductions and casual conversations about current events, as well as work, family, and autobiographical information; can handle limited work requirements, needing help in handling any complications or difficulties; can get the gist of most conversations on non-technical subjects (i.e., topics which require no specialized knowledge) and has a speaking vocabulary sufficient to express himself simply with some circumlocutions; accent, though often quite faulty, is intelligible; can usually handle elementary constructions quite accurately but does not have thorough or confident control of the grammar.

Level 3 - MINIMUM PROFESSIONAL PROFICIENCY. Able to speak the language with sufficient structural accuracy and vocabulary to participate effectively in most formal and informal conversations on practical, social, and professional topics. Can discuss particular interests and special fields of competence with reasonable ease; comprehension is quite complete for a normal rate of speech; vocabulary is broad enough that he rarely has to grope for a word; accent may be obviously foreign; control of grammar good; errors never interfere with understanding and rarely disturb the native speaker.

Level 4 - FULL PROFESSIONAL PROFICIENCY. Able to use the language fluently and accurately on all Levels normally pertinent to professional needs. Can understand and participate in any conversation within the range of his experience with a high degree of fluency and precision of vocabulary; would rarely be taken for a native speaker, but can respond appropriately even in unfamiliar situations; errors of pronunciation and grammar quite rare; can handle informal interpreting from and into the language.

Level 5 - NATIVE OR BILINGUAL PROFICIENCY. Speaking proficiency equivalent to that of an educated native speaker. Has complete fluency in the language such that his speech on all Levels is fully accepted by educated native speakers in all of its features, including breadth of vocabulary and idiom, colloquialisms, and pertinent cultural references.

Quite appropriately, the testing procedure adopted to place examinees at the proper level on this proficiency scale was a carefully structured face-to-face conversation between the examinee and a native speaker of the test language, the latter often accompanied by a second examiner highly trained in the elicitation and rating process, who would monitor the conversation to insure that an adequate depth and range of speaking opportunities had been provided. A well-conducted interview, although appearing on the surface to be a friendly and informal conversation on a variety of topics of interest to the examinee, was at the same time a carefully structured and orchestrated attempt on the testers' part to determine the examinee's ability or lack of ability to perform the language-use tasks at issue for each of the scale levels, and with the degree of facility and precision specified for that level. Interviewees who were solidly able to handle the performance requirements for a given scale level but not go appreciably beyond that level were awarded the level designation in question (1, 2, 3, etc.). Interviewees whose performance was appreciably superior to the base-line requirements for a given level but who failed to meet in all respects the requirements for the next higher level received an additional "plus" score (1+, 2+, 3+, etc.). Together with a "zero" point representing absolutely no functional ability in the language, the FSI proficiency scale had a range of 11 possible scores: 0, 0+, 1, 1+, 2, 2+, 3, 3+, 4, 4+ and 5.

Over a period of years, the "FSI interview" and its accompanying rating scale were used with a high degree of operational success and satisfaction by both the FSI and several other government agencies, including the Defense Language Institute (DLI), the Language School of the Central Intelligence Agency (CIA), and the U.S. Peace Corps. The latter agency played a considerable role in expanding the use of the FSI proficiency scale and interview as a result of a major project it undertook in the early 1970s to train cadres of local interviewer/raters at a large number of Peace Corps in- country language training sites worldwide. To assist in this very ambitious task, which soon began to strain its own staff resources, the Peace Corps called on the services of Educational Testing Service (ETS). ETS designed and conducted an "exportable" interviewer/rater training program in which ETS staff members, previously trained by FSI in the interviewing and rating process, traveled to designated Peace Corps in-country sites to instruct and certify local Peace Corps staff in the use of these techniques. Accompanying kits of training materials were developed to supplement the face-to-face training; these included workbooks and audiotapes of sample interviews in Spanish (for use in use in Central/Latin America training sites) and French (Francophone Africa). An additional English-medium training set was also developed to support interviewer/rater training in a number of other Peace Corps languages (e.g., Twi, Trukese, Nepalese, Korean). In these instances, the ETS trainer would provide instruction and interviewing/rating practice in English, which the in-country participants would subsequently carry over to their own operational interviewing and rating in the host country language.

The visibility and scope of the ETS/Peace Corps project contributed substantially to the development of more widespread interest in the "FSI" testing procedure, both within and outside the federal sector. A second major factor in the dissemination of this testing approach within the academic community was a series of three "Testing Kit Workshops" planned and implemented by Dr. James R. Frith, then dean of the FSI School of Language Studies. Liskin-Gasparro (1984) describes in detail the purpose and nature of these workshops, which provided widespread familiarization of college and university language teachers in the FSI interview and rating process. The Testing Kit project also resulted in the development of workbook/audiotape training materials in French and Spanish and in the quite successful implementation of a "distance" follow-up procedure following the initial training, in which prospective interviewer/raters would work with the kit materials, interview and rate students on their own campuses, and forward sample tapes of their interviews, together with the scores assigned, to FSI-trained workshop leaders for monitoring and critique.

A third major development and dissemination initiative within the same (early 1980s) time frame was the so-called "Common Yardstick" project conducted under a U.S. Department of Eduction grant to ETS. This project involved the convening of a group of academic and government agency representatives highly knowledgeable in guideline-based interviewing and scoring in both FSI and other contexts, to discuss the FSI Level descriptions (as well as additional verbally defined scales developed by the British Council and other European agencies) and to determine what modifications of the scale might make the testing procedure and the information derived more broadly useful across a wide range of academic and other applications. Among the recommendations of the Common

Yardstick participants was that the scale be made more "fine-grained" at the lower end of the proficiency range, that is, be revised to provide a greater number of distinctions between FSI Levels 0 and 2 than the five scores (0, 0+, 1, 1+, and 2) available within that range on the traditional FSI scale. This recommendation arose in large part from the recognition that, as a consequence of the extremely broad range of proficiency represented by the total 0-5 scale, the performance of students within most regular academic settings (including even college language majors; see Carroll, 1967) would not be expected to go beyond Level 2 or 2+. Given this restricted range of "possible" scores, it would be quite likely that, for example, a student rated at Level 1 at the end of a year of high school language study would not show further progress on the FSI scale over the next one or even two years in that, even though a variety of aspects of the language were indeed being acquired by the student on a more discrete-point basis, these quite respectable advances in achievement might not be sufficient to produce the overall proficiency required at Level 2 (or 1+) on the original scale.

In an effort to provide both a psychometrically and publicly acceptable solution to this restriction-of-range problem, the Common Yardstick project proposed a modified version of the original FSI scale, which was refined through further Department of Education grants to ETS and to the American Council on the Teaching of Foreign Languages (ACTFL) and subsequently disseminated within the profession as the "ACTFL/ETS Speaking Proficiency Guidelines." At Level 1+ and beyond on the traditional FSI scale, the FSI scale and the modified ACTFL/ETS guidelines reflect functionally equivalent performance levels, although with slight variations in the verbal descriptions. However, from FSI Levels 0 through 1, the ACTFL/ILR scale contains five separate performance levels, with associated descriptions, rather than the three Levels (0, 0+, 1) available in the FSI scale. In addition to providing more scoring categories at the lower end of the scale, the ACTFL/ETS guidelines make use of short verbal labels for each of the rating categories (e.g., Novice-High, Intermediate-Mid). These labels were intended to be more appealing from psychological and motivational perspectives than the FSI numerical values (e.g., 0+). The descriptive labels for each of the ACTFL/ETS scale values, together with, wherever applicable, the corresponding FSI designations, are shown below.

FSI	ACTFL/ETS
0+	Novice-High
1	Intermediate-Low
	Intermediate-Mid
1+	Intermediate-High
2	Advanced
2+	Advanced-Plus
3 and higher	Superior

The ACTFL/ETS scale makes no distinctions among proficiency levels higher than FSI Level 3, in recognition of the scarce need for measurement discriminations beyond this Level in most regular academic contexts.

For convenience, the ACTFL/ETS verbal descriptions for speaking performance are reproduced immediately below.

NOVICE. The Novice Level is characterized by an ability to communicate minimally with learned material.

Novice-Low. Oral production consists of isolated words and perhaps a few high-frequency phrases. Essentially no functional communicative ability.

Novice-Mid. Oral production continues to consist of isolated words and learned phrases within very predictable areas of need, although quantity is increased. Vocabulary is sufficient only for handling simple, elementary needs and expressing basic courtesies. Utterances rarely consist of more than two or three words and show frequent long pauses and repetition of interlocutor's words. Speaker may have some difficulty producing even the simplest utterances. Some Novice-Mid speakers will be understood only with great difficulty.

Novice-High. Able to satisfy partially the requirements of basic communicative exchanges by relying heavily on learned utterances but occasionally expanding these through simple recombinations of their elements. Can ask questions or make statements involving learned material. Shows signs of spontaneity, although this falls short of real autonomy of expression. Speech continues to consist of learned utterances rather than of personalized, situationally adapted ones. Vocabulary centers on areas such as basic objects, places, and most common kinship terms. Pronunciation may still be strongly influenced by first language. Errors are frequent and, in spite of repetition, some Novice-High speakers will have difficulty being understood even by sympathetic interlocutors.

INTERMEDIATE. The Intermediate Level is characterized by an ability to create with the language by combining and recombining learned elements, though primarily in a reactive mode; initiate, minimally sustain, and close in a simple way basic communicative tasks; and ask and answer questions.

Intermediate-Low. Able to handle successfully a limited number of interactive, task-oriented and social situations. Can ask and answer questions, initiate and respond to simple statements, and maintain face-to-face conversation, although in a highly restricted manner and with much linguistic inaccuracy. Within these limitations, can perform such tasks as introducing self, ordering a meal, asking directions, and making purchases. Vocabulary is adequate to express only the most elementary needs. Strong interference from native language may occur.

Intermediate-Mid. Able to handle successfully a variety of uncomplicated, basic and communicative tasks and social situations. Can talk simply about self and family members. Can ask and answer questions and participate in simple conversations on topics beyond the most immediate needs; e.g., personal history and leisure-time activities. Utterance length increases slightly, but speech may continue to be characterized by frequent long pauses, since the smooth incorporation of even basic conversational strategies is often hindered as the speaker struggles to create appropriate language forms. Pronunciation may continue to be strongly influenced by first language and fluency may still be strained. Although misunderstandings still arise, the Intermediate-Mid speaker can generally be understood by sympathetic interlocutors.

Intermediate-High. Able to handle successfully most uncomplicated communicative tasks and social situations. Can initiate, sustain, and close a general conversation with a number of strategies appropriate to a range of circumstances, but errors are evident. Limited vocabulary still necessitates hesitation and may bring about slightly unexpected circumlocution. There is emerging evidence of connected discourse, particularly for simple narration and/or description. The Intermediate-High speaker can generally be understood even by interlocutors not accustomed to dealing with speaking at this Level, but repetition may still be required.

ADVANCED. The Advanced Level is characterized by an ability to converse in a clearly participatory fashion; initiate, sustain, and bring to closure a wide variety of *communicative tasks, including those that require an increased ability to convey meaning with diverse language strategies due to a complication or an unforeseen turn of events; satisfy the requirements of school and work situations; and narrate and describe with paragraph-length connected discourse.*

Advanced. Able to satisfy the requirements of everyday situations and routine school and work requirements. Can handle with confidence but not with facility complicated tasks and social situations, such as elaborating, complaining, and apologizing. Can narrate and describe with some details, linking sentences together smoothly. Can communicate facts and talk casually about topics of current public and personal interest, using general vocabulary. Shortcomings can often be smoothed over by communicative strategies, such as pause fillers, stalling devices, and different rates of speech. Circumlocution which arises from vocabulary or syntactic limitations very often is quite successful, though some groping for words may still be evident. The Advanced-Level speaker can be understood without difficulty by native interlocutors.

Advanced-Plus. Able to satisfy the requirements of a broad variety of everyday, school, and work situations. Can discuss concrete topics relating to particular interests and special fields of competence. There is emerging evidence of ability to support opinions, explain in detail, and hypothesize. The Advanced-Plus speaker often shows a well-developed ability to compensate for an imperfect grasp of some forms with confident use of communicative strategies such as paraphrasing and circumlocution. Differentiated vocabulary and intonation are effectively used to communicate fine shades of meaning. The Advanced-Plus speaker often shows remarkable fluency and ease of speech, but under the demands of Superior Level, complex tasks, language may break down or prove inadequate.

SUPERIOR. The Superior Level is characterized by an ability to participate effectively in most formal and informal conversations on practical, social, professional, and abstract topics; and --support opinions and hypothesize using native-like discourse strategies.

Superior. Able to speak the language with sufficient accuracy to participate effectively in most formal and informal conversations on practical, social, professional, and abstract topics. Can discuss special fields of competence and interest with ease. Can support opinions and hypothesize, but may not be able to tailor language to audience or discuss in depth highly abstract or unfamiliar topics. Usually the Superior Level speaker is only partially familiar with regional or other dialectical variants. The Superior Level speaker commands a wide variety of interactive strategies and shows good awareness of discourse strategies. The latter involves the ability to distinguish main ideas from supporting information through syntactic, lexical, and suprasegmental features (pitch, stress, intonation). Sporadic errors may occur, particularly in low-frequency structures and some complex high-frequency structures more common to formal writing, but no patterns of error are evident. Errors do not disturb the native speaker or interfere with communication.

Roughly simultaneously with the elaboration of the ACTFL/ETS Proficiency Guidelines for speaking, these two organizations, in collaboration with representatives from DLI, FSI, NSA, and other government language teaching agencies as well as the academic community, prepared draft proficiency guidelines for listening comprehension, reading comprehension, and writing. For reading comprehension, level descriptions that had been developed by FSI as an additional component of its mid-1950s project were used as a point of departure; for listening comprehension and writing, the working groups were required to operate virtually from scratch. The final guideline descriptions for all four skill areas, incorporating modifications based on widespread dissemination and review of earlier versions through ACTFL and other channels, are provided in Byrnes and Canale (1987) and are also available directly from ACTFL.

From about 1982 until quite recently, ACTFL has worked with committees of secondary school and college French, German, and Spanish teachers to develop language-specific proficiency guidelines for each of the four skill areas in these three languages. In addition, ACTFL-coordinated guidelines for Chinese, Japanese, and Russian are nearing completion, and guidelines in Indonesian, Hindi, and Swahili are under development. These language-specific guidelines, based in all instances on the generic guidelines, are intended to provide reasonably detailed information concerning the specific lexical, structural, and phonological features of that language that are usually associated with performance at a particular guideline level.

Development of elicitation and scoring procedures – In addition to development and dissemination of the verbal proficiency guidelines themselves, evolutionary and in some instances more radical changes have taken place since the original FSI initiatives, with respect to both the procedures used to elicit examinee performance and to assign proficiency level scores based on this performance. As discussed in Wilds (1975), the interviewing procedure originally adopted by FSI typically made use of two separate testers: a native speaker of the test language and a certified language examiner, who could be either an experienced native instructor in that language or a linguist thoroughly familiar with the language. The native speaker tester was responsible for conducting the bulk of the interview, by engaging the examinee in a structured conversation of suitable length (approximately 10 to 40 minutes depending on examinee proficiency level), during which time the linguist would sit to the side and take unobtrusive notes on particular aspects of the examinee's performance, occasionally entering the conversation to probe the examinee's command of a particular structural or other feature of the language that the primary interviewer had neglected or not sufficiently explored. Toward the end of the interview, a communicative "situation" would often be set up, in which one of the testers would take the role of a monolingual speaker of the test language and the other, a monolingual speaker of English. The examinee's task in this situation was to serve as an informal interpreter between the two testers; depending on the proficiency level of the examinee, the situation presented could range from a relatively simple "survival" task (e.g., making a hotel reservation) to a considerably more challenging requirement (remodeling an apartment or negotiating a lease).

With regard to rating procedure, the note-taking member of the interviewing team was usually considered the primary authority for making the initial level judgment, since he or she was usually in a better position than the "conversing" partner to carefully and critically evaluate the examinee's performance over the course of the interview. Consultation between the two interviewers was, however, encouraged, and the level rating finally assigned was usually a matter of full consensus. The process used to arrive at a rating, whether consciously or implicitly, was to serially determine whether or not the examinee had met in all respects the stated performance requirements of each successive level, up to the point at which all the requirements of a given level had *not* been met. The remaining decision was whether the examinee had "substantially exceeded" typical performance expectations of the immediately lower level (in which case a "plus" value would be added), or had simply met in full the lower level requirements, thus warranting the basic level rating. Although the official verbal level definitions were considered authoritative for rating purposes, an evaluation aid in the form of a two-way grid of linguistic "factors" to be considered in arriving at a rating was developed. This grid showed listening comprehension, accent, grammar, vocabulary, and fluency along one dimension and the five major level categories (1, 2, 3, 4, 5) along the other, with each cell of the grid containing a short verbal description of anticipated examinee performance for that linguistic factor at that level. Performance patterns, as shown on the grid, were of assistance in assigning "plus" values where merited; they also served to emphasize the "non-compensatory" nature of the rating process in that, regardless of relatively higher examinee performance on one or even several of the individual factors, the final global rating, by definition, was restricted to the highest level at which the requirements of *all* factors had been satisfied (with the possibility of an additional "plus" as appropriate). Experience with the rating scale and the grid definitions frequently pointed to grammatical control as the "weak link" in examinee performance, with highly fluent speakers with extensive vocabulary, a high level of listening comprehension, and accurate pronunciation unable to successfully accomplish Level 3 tasks. Elicitation and scoring procedures initially used by other government agencies followed those of the FSI for the most part, with some minor differences. For example, testing at CIA was normally conducted by two native speaker instructors, with each tester sharing the interviewing role more equally than in the typical FSI interviewer/linguist-notetaker approach.

Through the efforts of Lowe and Clifford at CIA and others, considerable additional structure and objectivity were provided both the interviewing and rating process in the form of tester training workshops and accompanying training materials, as well as through increased professional discussion and joint development efforts among agencies and institutions involved in oral proficiency testing. A significant conceptual advance, with major implications for both elicitation and rating procedures, was the elaboration of the so-called "Functional Trisection" (Higgs & Clifford, 1982), in which the performance requirements at each of the proficiency Levels were analyzed as consisting of three separate components: function, content, and accuracy. "Function" refers to the types of language-use tasks the speaker is expected to be able to carry out at the specified level. For example, the Level 1 speaker must be able to "create" with the language by asking and answering factual questions which combine and recombine previously learned language elements in new ways for genuine communicative purposes. A brief description of Level 2 functions includes the examinee's ability to "fully participate in casual conversations, [express] facts, give instructions, describe, report on, and provide narration about current, past, and future activities." (Higgs & Clifford, 1982, p. 63.) "Content" refers to the kinds of topics or subject areas at issue in the communication. For example, a Level 2 speaker is expected to be able to speak about "concrete topics, such as own background, family, and interests, work, travel, and current events." A Level 3 speaker, by comparison, has a considerably wider content repertoire including the ability to communicate on "practical, social, professional and abstract topics, particular interests, and special fields of competence." The third element of the trisection, "accuracy," describes the degree of structural, lexical, and phonological precision with which the examinee is able to communicate, and also includes aspects of discourse competence as these affect the naturalness, efficiency, and overall acceptability of the communication to native speakers. The accuracy requirements for Level 3 and Level 4 performance are characterized as follows: (Level 3) "Errors never interfere with understanding and rarely disturb the native speaker. Only sporadic errors in basic structures." (Level 4) "Nearly equivalent to an educated native speaker. Speech is extensive, precise, appropriate to every occasion, with only occasional errors."

Lowe (1982) has codified, as well as conducted numerous training workshops on, the basic types of elicitation procedures involved in a well-conducted interview. The entire interview is viewed as consisting of four separate phases: a "warm-up," occupying the first few minutes of the interview, during which the tester attempts to put the examinee at ease as well as to obtain a rough indication of his or her general ability level; a series of "probes," in which the pace and level of sophistication of the discourse are intentionally raised in order to determine the examinee's ability or lack of ability to perform at a higher level; "level checks," interspersed with the probes, in which the intent is to confirm the examinee's firm possession of a given proficiency level by broadening the base of content and topics discussed at that level; and, at the end of the interview, a "wind-down" period in which, to provide a feeling of closure and "success" on the examinee's part, the conversation is returned to a level that the examinee is known to be able to handle readily and effectively.

Although some variations continue to exist in the interviewing procedures used at different government language agencies --especially with regard to a recently revised elicitation process at FSI which makes use of, among other procedures, "briefing" monologues on the examinee's part in addition to the traditional conversational interaction --current government and academic community testing procedures generally make use of the basic interview phases outlined by Lowe, as well as incorporate the functional trisection concept, either explicitly or implicitly, in the elicitation and rating process. Also in reasonably frequent use, especially by ACTFL-trained interviewers, are printed "situation cards" which provide a brief description in English of a real-life setting in which the examinee is required to carry out a specified communication task (for example, telephone a garage to report an automobile breakdown and arrange for a tow truck), with the interviewer playing the role of the native speaker. Situation cards have been developed for each of the ACTFL scale levels, tailored to reflect the function and content requirements at issue at the given level.

Mention should also be made of elicitation procedures for speaking proficiency testing which do not involve a live interviewer or other interlocutor, but instead make use of a test booklet and/or audiotape to present the stimuli to which the examinee responds. These procedures, collectively referred to as "semi-direct" (Clark, 1979), date back to at least the early 1960s in the form of the tape- and booklet-mediated MLA-Cooperative Tests in French, German, Italian, Russian, and Spanish developed by ETS under contract with the U.S. Office of Education (ETS, 1961). Although eliciting actual speech on the part of the examinee, the MLA-Cooperative Tests were not scored on a proficiency-scale basis but on cumulative points awarded for each of several discrete speaking activities (e.g., sentence repetition, reading aloud, description of pictures). More recently, DLI has used a tape-and-booklet format for semi-direct speaking testing in several languages under its Defense Language Proficiency Test (DLPT) program. Although there is general correspondence in some of the elicitation procedures for the MLA-Cooperative and DLPT speaking tests (for example, both ask the examinee to give narrative descriptions of pictured scenes or activities), the DLPT is scored holistically in accordance with the ILR verbal descriptions. Also under government auspices, Lowe and Clifford (1980) developed the Recorded Oral Proficiency Examination (ROPE). The ROPE is also evaluated in terms of the ILR guidelines, but uses only a stimulus audiotape (with no accompanying booklet) on which conversational questions of increasing complexity and sophistication are posed by a native speaker. Within the academic community, tape-and-booklet-mediated semi-direct tests have recently been developed in Japanese and Chinese at CAL and in Hindi at the University of Pennsylvania; all of these tests are scored according to the ACTFL proficiency scale. Although the elicitation procedures for semi-direct tests do not permit true interactive conversation, they have consistently been found to correlate very highly with direct proficiency interviews (Lowe & Clifford, 1980; Clark, 1986a), and on this basis are considered to be viable alternatives to direct interviewing in situations where the latter approach is not practically or economically feasible.

Proficiency-based testing of skills other than speaking – Notwith-standing the historically greater public visibility of the proficiency guidelines and testing procedures used in assessing speaking, a substantial amount of scale elaboration and test development has taken place for the other skill areas as well. As previously indicated, concurrent with the initial development of speaking proficiency guidelines at FSI was the preparation of a parallel set of verbal descriptions of five levels of reading proficiency, ranging from basic "survival" tasks (e.g., understanding common street signs, shop designations, etc.) to the full reading comprehension abilities of an educated native speaker. The FSI elicitation procedure for reading --which was usually tested at the same time as speaking --was to have the examiner, on a face-to-face basis, present a series of graded texts to the examinee for oral translation into English. Listening comprehension was not formally tested as a separate skill, but was estimated in a general way in the course of the speaking proficiency interview. Writing ability, a minor component of the functional language needs of foreign service personnel, was not routinely tested at FSI, nor has there been much emphasis on the teaching or testing of this skill in the great majority of other government and academic programs.

Through the joint efforts of member agencies of the Interagency Language Roundtable (ILR), detailed verbal descriptions of "survival"-to-native levels of proficiency in listening comprehension and reading (as well as in writing, for the sake of completeness) were developed and adopted by these agencies in 1981, together with the previously elaborated speaking scale, as representing the official agency-wide proficiency standards for all four skill areas. On the academic side and in close coordination with the ILR, ACTFL coordinated the development of the previously-mentioned "fine-grained" level descriptions for listening comprehension, reading, and writing, incorporating the same "Novice," "Intermediate" and "Advanced" (plus sub-categories) format used for the speaking scale. With respect to testing procedures for listening comprehension and reading, CIA, DLI, and NSA have developed tape-and-booklet-based objective tests in a large number of target languages. These tests use authentic spoken material and printed texts representative of the various proficiency scale Levels, and are scored and interpreted according to the ILR guidelines. On the academic front, ETS has developed a proficiency-based listening and reading test in Japanese covering the Level 1 to Level 3+ range and is completing a similar test in Russian. The Center for Applied Lingustics has available for operational use a listening/reading proficiency test in Chinese covering a similar proficiency range, and the University of Pennsylvania has recently completed an a parallel test development project for Hindi.

2. Current status

Taken as a whole, proficiency-based testing of the active speaking and receptive (listening/reading) skills constitutes at the present time a very large-scale undertaking across both government and academic communities. During 1986, an estimated 15,000 language proficiency tests were administered by ILR member agencies. Of these, about 4,000 tests were administered by FSI and covered speaking (including "interactive" listening) and reading. Almost all of the remaining tests assessed candidates' performance in each of the three skill areas of listening, speaking, and reading. ACTFL has moved strongly to establish a solid infrastructure for speaking testing based on the "expanded" proficiency guidelines and has currently trained and certified over 300 interviewer/raters nationwide. Although the total number of official interviews conducted by ACTFL-trained testers is still reasonably low (about 200 in 1986), it is expected that this figure will grow considerably over the near-term future. Anticipated sources of impetus include an increase in the number of states requiring demonstrated speaking proficiency for teacher certification (such requirements have already been established by California, Florida, Illinois, New Jersey, and Texas), as well as growing interest on the part of both government agencies and the business sector in a testing system that allows job candidates to demonstrate their language competence in a standardized and easily interpreted way.

In addition to the "official" interviews conducted and formally reported by certified testers, a quite large number of unofficial speaking tests are being conducted annually in secondary school and college settings both by certified testers and by teachers and supervisors who have had at least some instructional exposure to the proficiency guidelines and elicitation techniques through presentations and workshops by ACTFL and other agencies. Although space does not permit discussion of the direct and major materials preparation and instructional design implications of proficiency-based testing (see Omaggio, 1984; Galloway, 1987), it may be expected that as more and more textbook publishers and program planners orient their efforts along proficiency development lines, teachers in these programs will find it increasingly important for them to become acquainted with and involved in proficiency-based testing issues.

3. Needed Research

The current and increasingly widespread use of the oral proficiency interview and proficiency-based tests for other skill areas, both in government language training programs and throughout the academic community, raises an imperative need for a well-planned and carefully coordinated program of research on critical theoretical and technical issues associated with these measurement procedures. The items outlined below may be considered indicative, but by no means exhaustive, of the variety of issues that could be very profitably addressed within a comprehensive program of research and development in the "proficiency testing" area.

1. Development of criterion measures external to the proficiency tests per se – Somewhat paradoxically, although the oral proficiency interview, as well as proficiency-based listening comprehension and reading tests, discussed earlier in this paper are considered to be more closely reflective of realistic language-use tasks than are, for example, the traditional discrete-element, achievement-based measures in earlier common use, these proficiency tests themselves do not correspond precisely, either in manner of administration or in range and depth of coverage, to the language-use situations that the examinee would in fact encounter in the real-life setting. Numerous examples of at least some degree of "shortfall" may be cited between real-life language use on the one hand and the types of tasks presented in "proficiency" tests on the other. For example, within the oral interview, the dialogue setting between examinee and tester necessarily differs in a number of respects from conversations that would arise in natural communicative settings, including overall communicative purpose, identity, status, and role of interlocutors, degree of availability of real-life communicative aids and prompts (e.g., menus that can be pointed to in a real-life restaurant), and so forth. In listening comprehension testing, it is for all practical purposes not possible to duplicate the real-world audio stimuli precisely: tape recorded material differs in numerous acoustic respects from "live"; environmental conditions and cues cannot be closely replicated, etc. In reading testing, although there is an appreciably greater opportunity for realism in stimulus presentation than is the case for speaking and listening, concerns about overall test length and examinee fatigue usually restrict the amount of material presented in a given test to well below that necessary to cover the wide variety of text types and topical areas implicit in, say, the Level 3 proficiency descriptions.

It would probably be fair to say that, although the communicative realism of current proficiency-based testing procedures is, by and large, substantially greater than that of more traditional measurement approaches, the admitted differences that continue to exist between proficiency tests and the real world raise genuine --and eminently researchable --questions about the type and degree of measurement loss that may be taking place. One possible approach, discussed in greater detail elsewhere (Clark, 1980; Bachman & Clark, in press) would be to develop series of quite lengthy, highly face- and content-valid testing exercises, each of which would approximate as closely as possible a particular real-life language use task within a given skill modality, and then correlate these results with those provided by the general proficiency tests. For example, a highly realistic criterial test might be developed and administered to determine the examinee's ability to read a daily newspaper with the same comprehension goals and results as a typical target language native (e.g., gain an overview knowledge of the main news items, check on the weather forecast, properly appreciate and interpret the day's editorial, enjoy the style and humor of certain favorite comic strips, etc.). Similarly functioning criterial tests of other aspects of the total "reading spectrum" that are either stated in or strongly implied by the proficiency level descriptions would also be prepared and administered. The observed relationships between examinee performance on these more detailed tests and on the general proficiency measure would permit considerably more specific and more meaningful statements about the types and extent of "reading" ability represented by a given level score on the latter instrument than is presently the case.

2. Development of mechanisms for gathering and reporting use-validity information in a variety of contexts – For purposes of information and decision making in real-life contexts, most operational users of proficiency-scale based test results (other than language teachers and researchers) are not so much interested in the detailed level descriptions per se as they are in knowing what a given score level says about the examinee's ability or lack of ability to carry out effectively the language-use tasks at issue in a particular type of employment or other specific "setting" for which that individual is being considered. For example, municipalities considering hiring non-native speakers of English for particular types of local service (firefighter, police) would want to know the minimum speaking proficiency level --expressed as a single number, *not* a detailed description --that would be needed to perform adequately the language-use tasks involved in that particular position. There are at least two avenues available for developing and disseminating this type of information. The first would be to conduct a detailed task analysis of the position with respect to its linguistic requirements, develop highly face- and content-valid measures of these particular tasks, and use this information to correlate the general proficiency levels with success in task accomplishment. Although this has in fact been done for some government positions involving substantial language requirements, the overall cost and complexity of this approach would proably rule it out of consideration in most instances. However, an alternative approach may be feasible in a number of settings: the correlation of measured proficiency levels with expert judgments, by supervisors or others in a position make these judgments, of the degree of communicative success exhibited by given individuals in the particular employment or other real-life settings at issue. Although reliability of the expert judgments would be one concern here, a variety of techniques could be used to increase the objectivity of the process, including, for example, asking the judges to rate the subjects with respect to each of several sub-components of performance, rather than on a single holistic scale. If this type of use- validity information could be obtained and reported for a fairly large number of different employment and other use settings, the proficiency scale levels would, over time, acquire even more extensive meaning and practical utility than they presently enjoy.

3. Further examination of the principles on which level descriptions are developed – As has been pointed out in a number of sources (Clark, 1972; Higgs & Clifford, 1982), the FSI proficiency level definitions (as well as the ACTFL adaptations) combine, within a single descriptive statement, both "linguistic" and "functional" components. For example, the FSI Level 2 description for speaking is for the most part functionally-oriented ("Can handle with confidence but not with facility most social situations, including introductions and casual conversations about current events, as well as work, family, and autobiographical information...."), but at the same time includes statements explicitly addressed to formal linguistic characteristics of the performance (e.g., "accent, though often quite faulty, is intelligible; can usually handle elementary constructions quite accurately...."). This may be attributable in large part to the fact that both the original FSI descriptions and their successors have attempted to serve two purposes simultaneously: first, inform the external "user" of the descriptions about the general types of real-life settings/language-use situations in which the examinee would be considered able to function; and second, provide sufficiently detailed guidance to the tester to allow him or her to properly elicit and rate the examinee's performance on the descriptive scale. It may be suggested that mixing both of these types of description within a single paragraph statement tends to reduce, at least to some extent, the potential accuracy and effectiveness of both kinds of information.

The previously-described "Functional Trisection" for speaking proficiency levels appears to be a step in the right direction, in that it breaks out linguistic accuracy (including discourse competence considerations) from both function and topic/content. A logical follow-up might be to explicitly separate these elements into two different descriptions for each proficiency level: an "interviewer/rater's description," setting forth in more detail than is presently the case the structural, lexical, phonological, and discourse management features associated with that level; and a "score user's" description, concentrating on the functional/topical/content capabilities of an individual rated at that level. Some of the additional performance aspects recently proposed by Bachman (Bachman & Clark, in press) as necessary components of a thoroughgoing model of communicative proficiency might be considered as candidates for inclusion in the "interviewer/rater" description; and the results of further use-validity studies of the type outlined in item (2) above could provide very relevant data for verifying and perhaps considerably expanding the description of functions and topic/content capabilities associated with each proficiency level.

4. Investigation of the relationship between native-speaker judgments of proficiency and assigned proficiency levels – Among the many hypotheses for investigation that a detailed construct validity examination of the speaking proficiency scale and rating technique would be expected to give rise to, one of these would surely be that the obtained ratings for given examinees are generally consistent with the judgments of educated native speakers not *a priori* familiar with this assessment approach. Although judgments by untrained persons made on adjectivally defined scales ("excellent," "good," etc.) can be highly unreliable, alternative approaches, such as repetitive forced choice between pairs of examinee performance samples, could be used to secure the native speaker judgments in a straightforward and psychometrically appropriate way. Particularly of interest in such a study would be the relationship between individual sub-elements of the total interview-based rating (e.g., the "factor" scores of listening comprehension, structure, fluency, pronunciation, and vocabulary) and native speaker judgments. Two other highly intriguing questions would be (a) whether or not the native speaker judgments would tend to corroborate the frequently cited "quantum leap" between Level 2+ and 3 with respect to range and effectiveness of speaking performance, and (b) whether native speakers would under naturalistic conditions actually make distinctions as fine-grained as those represented by adjacent sub-levels of the ACTFL scale (e.g., "Novice-Mid" vs. "Novice-High").

5. Detailed investigation of the scoring reliability of both ILR and ACTFL guideline-based speaking tests – Guideline-based speaking proficiency testing presents a veritable cornucopia of scoring reliability questions and associated research opportunities. For example, with regard to "direct" (interview-based) testing, a major practical concern is the length of time the initial rater training may be considered to "last" before unacceptable scoring variations begin to occur. Intra-rater reliability studies based on the repetitive scoring of a set of fixed interviews over varying time periods could address this question, and could also help to determine optimum intervals for retraining. In such an investigation, an important variable that should be explicitly provided for is the amount and distribution of the individual's "operational" testing work (i.e., testing/rating activities other than those associated with the study) as this affects the maintenance of or decrease in scoring reliability over time.

A second major consideration is inter-rater reliability, not only among members of a particular testing "unit" (for example, French teacher/testers in a high school language department, Russian testers at DLI), but also on a more widespread geographical and inter-organizational basis. A recent study by Clark (1986b) showed appreciable variation in the proficiency levels assigned to a common group of examinees when tested by interviewer/raters from three government language training agencies, an outcome suggesting that testing groups that are largely "self-contained" with respect to both interviewer/rater training and operational testing may develop idiosyncratic rating tendencies by comparison to testers in other organizations. The inter-rater reliability of ACTFL-trained testers, working largely autonomously in secondary school and undergraduate settings, also warrants detailed examination. The need to carry out a wide variety of inter-rater reliabiltiy studies across a number of different organizations and testing situations suggests very strongly the advisability of developing and making available comprehensive sets of "master" tapes of representative interviews at each score level, to be used as a common corpus for rating by participants in these studies. Such a series of tapes could also serve as an "institutional memory" within or across organizations, in that they could be re-scored periodically to determine the direction and amount of any scoring drift that might have occurred over time.

Test-retest reliability of the interviewing/rating process is also an important area of inquiry, since in the absence of measurement stability over relatively short time spans (periods in which no true proficiency change would be expected to have taken place), examinees tested on any one occasion could claim, with considerable justification, that their rating was primarily a function of the time of day, the mood of the interviewer, the particular questions asked, or numerous other happenstance conditions rather than truly reflective of their actual proficiency. The only effective and psychometrically valid response to these concerns would be the empirical demonstration of a satisfactory degree of "robustness" and invariance of interview scores across a variety of test-retest conditions.

Many of the relability issues that arise in direct proficiency interviewing and rating are also relevant to semi-direct testing. One potential source of unreliability in interview-based testing --variation in test content --is of course not present in the semi-direct mode; however, intra- and inter-rater reliability, as well as test-retest reliability, still merit close examination. A small-scale (two rater) reliability study of the semi-direct Chinese Speaking Test recently developed by CAL found linear correlations in the low to mid 90s for inter-rater reliability and generally similar figures for test-retest (alternate form) reliabilities for both same-rater and different-rater conditions (Clark, 1986a). However, a fair degree of variation in the absolute values of the assigned ratings was observed, suggesting that close attention to the proper initial "anchoring" of the raters' judgments, as well as periodic scoring congruence checks, would be needed to minimize this phenomenon.

Although these and other areas of research and development may be cited as desirable objectives for ongoing and future efforts in the proficiency testing field, this overview of potential additional activities should not be concluded without an explicit acknowledgement of the enormous amount of solid research-based and practical information that has been accumulated over the past thirty years, and especially within the last eight to ten years, on the appropriate use and interpretation of proficiency-based testing techniques. In this regard, the author would request the reader's indulgence in reproducing, with only slight adaptation, the concluding portion of a related paper given some twelve years ago (Clark, 1975), the main emphasis of which is no less true or applicable today.

All in all, how should tests such as the FSI interview and other procedures aimed at determining "proficiency" finally be evaluated? Do the research needs that have been identified mean that current techniques are not at all useful, or that we are defeated before we begin in our efforts to measure our students' language proficiency against real-life standards? An interesting parallel might be drawn between this situation and the scientific investigation of the bumblebee which, if my information is correct, was carried out a few years back by an aerodynamic engineer. This specialist carefully measured the wingspan, body weight, airflow pattern, size and placement of wing muscles, number of wing beats per second, and numerous other aspects of the bee's physiology, and by means of elaborate diagrams, mathematical formulas, and computer-aided calculations, was able to demonstrate conclusively that the bumblebee is incapable of flight.

Some of the observations made in this paper should perhaps be taken with the same grain of salt as the findings of the aerodynamic specialist, because it is undeniable that useful "proficiency testing" is indeed taking place at the present time by means of the FSI interview and various other techniques, and results are being obtained which both the test developers and the users of the test results are finding of considerable meaning and practical value. On the other hand, the fact that the "proficiency testing bumblebee" is indeed, getting through the air at the moment should not deter us from continuing to study, discuss, experiment, and through continued genetic engineering, develop improved kinds of proficiency testing procedures that will fly faster and farther and have an even greater useful payload.

References

Bachman, L. P., & Clark, L. D. (in press). The measurement of foreign/second language proficiency. *In The Annals of the American Society of Political and Social Sciences.*

Byrnes, H., & Canale, M. (Eds.). (1987). *Defining and developing proficiency: Guidelines, implementations, and concepts.* Lincolnwood, IL: National Textbook Company.

Carroll, J. B. (1967). *The foreign language attainments of language majors in the senior year: A survey conducted in U.S. colleges and universities.* Cambridge, MA: Harvard University Graduate School of Education.

Clark, J. L. D. (1972). *Foreign language testing: Theory and practice.* Philadelphia, PA: Chilton/Center for Curriculum Development.

Clark, J. L. D. (1975). Testing the foreign language skills for mastery. In R. McConnell & A. Papalia (Eds.), *Papers Presented at the Fourth International Conference of the Ontario Modern Language Teachers' Association and the New York State Association of Foreign Language Teachers* (pp. 50-57).

Clark, J. L. D. (1979). Direct vs. semi-direct tests of speaking ability. In E. J. Brière & F. B. Hinofotis (Eds.), *Concepts in language testing: Some recent studies* (pp. 35-49). Washington, DC: Teachers of English to Speakers of Other Languages.

Clark, J. L. D. (1980). Toward a common measure of speaking proficiency. In J. R. Frith (Ed.), *Measuring spoken language proficiency* (pp. 15-26). Washington, DC: Georgetown University Press.

Clark, J. L. D. (1986a). *Development, validation, and dissemination of a proficiency-based test of speaking ability in Chinese and an associated assessment model for other less commonly taught languages* (Final Project Report for Grant No. G008402258, U. S. Department of Education). Washington, DC: Center for Applied Linguistics.

Clark, J. L. D. (1986b). *A study of the comparability of speaking proficiency interview ratings across three government language training agencies.* Washington, DC: Center for Applied Linguistics.

Clark, J. L. D. (1961). *MLA-Cooperative Foreign Language Proficiency Tests.* Princeton, NJ: Educational Testing Service.

Galloway, V. (1987). From defining to developing proficiency: A look at the decisions. In H. Byrnes & M. Canale (Eds.), *Defining and developing proficiency: Guidelines, implementations, concepts* (pp. 25-73). Lincolnwood, IL: National Textbook Company.

Higgs, T. V., & Clifford, R. T. (1982). The push toward communication. In T. V. Higgs (Ed.), *Curriculum, competence, and the foreign language teacher* (pp. 57-79). Skokie, IL: National Textbook Company.

Liskin-Gasparro, J. E. (1984). The ACTFL proficiency guidelines: A historical perspective. In T. V. Higgs (Ed.), *Teaching for proficiency, the organizing principle* (pp. 11-42). Lincolnwood, IL: National Textbook Company.

Lowe, P., Jr. (1982). *Manual for LS Oral Interview Workshop.* Washington, DC: DLI/LS Joint Oral Interview Transfer Project.

Lowe, P., Jr., & Clifford, R. T. (1980). Developing an indirect measure of overall oral proficiency. In J. R. Frith (Ed.), *Measuring spoken language proficiency* (pp. 31-39). Washington, DC: Georgetown University Press.

Omaggio, A. C. (1984). The proficiency-oriented classroom. In T. V. Higgs (Ed.), *Teaching for proficiency, the organizing principle* (pp. 43-84). Lincolnwood, IL: National Textbook Company.

Sollenberger, H. E. (1978). Development and current use of the FSI oral interview test. In J. L. D. Clark (Ed.), *Direct testing of speaking proficiency: Theory and application* (pp. 1-12). Princeton, NJ: Educational Testing Service.

Wilds, C. P. (1975). The oral interview test. In R. L. Jones & B. Spolsky (Eds.), *Testing language proficiency* (pp. 29-44). Arlington, VA: Center for Applied Linguistics.

THE ACTFL PROFICIENCY GUIDELINES: AN UPDATE

Judith E. Liskin-Gasparro
Educational Testing Service and Middlebury College

I have been asked for this symposium to provide an update on the ACTFL Proficiency Guidelines, specifically to comment on two areas: (1) the changes between the 1982 provisional guidelines and the revised 1986 version; (2) the implications of the guidelines for instruction. I have chosen to address, albeit briefly, curriculum design and classroom testing, leaving the question of teacher training for a paper later in this conference.

1. Introductory comments

Before I discuss the changes between the provisional and the revised guidelines, I would like to begin with some introductory comments. For background information on the ACTFL Guidelines, specifically the national trends that resulted in the drafting of the original, "provisional" version in 1982, I would refer you to David Hiple's (1986) chapter in the most recent volume of the ACTFL Foreign Language Education Series. Basically, the provisional guidelines were the result of a concatenation of events both inside and outside the foreign language profession: notably, a recommendation by the MLA Task Force on the Commonly Taught Languages, a mandate from President Carter's Commission on Foreign Language and International Studies, and the Educational Testing Service's Common Yardstick Project, funded by the US Department of Education. All of these were concerned with the question of establishing standards for the attainment of skill in a foreign language, standards that would go beyond the confines of course content and years of study to embrace notions of general linguistic competence, communicative skill, and the ability to use the foreign language in real-life settings beyond the classroom.

In addition, the climate for these recommendations and for the Common Yardstick Project had been informed by a decade of growing familiarity with the method used by the US government, particularly the Department of State, to assess the foreign language oral proficiency of its staff. Largely through the involvement of ETS personnel in the oral proficiency testing of Peace Corps volunteers, the interview procedure and the scale became known outside the government and were used in a limited number of civilian contexts during the 1970's --bilingual teacher certification in several states, the Experiment in International Living, and the Missionary Training Center of the Church of Jesus Christ of Latter Day Saints. The FSI Testing Kit workshops, held in 1979 and 1980, represented a more direct attempt to explore the possible applications of the oral proficiency interview and rating scale in post-secondary institutions.

These relatively limited but close encounters by academia with the oral proficiency interview in the late '70's and early '80's engendered a good deal of interest. There was a growing sense in the foreign language field that indeed the time had come to measure performance independent of a circumscribed program of study. Perhaps this was due in part to the national trend toward pre-professionalism in undergraduate education, and the corresponding expansion of foreign language department course offerings beyond the traditional literature program to include advanced-level course work in language as an auxiliary skill for other fields, such as business, the health professions, and social services.

A second, and related reason for interest was the model that the oral interview presented for testing the speaking skill. Until that time, the only tests of speaking available nationally were the MLA Cooperative Tests, long outdated, and the College Board Advanced Placement Examination in French Language, which was given only to a smaller number of advanced students at the high school level, and was not available for institutional use.

Finally, there was a trend, paralleled in the field of English composition, towards global evaluation of the productive skills. The 1970's saw the inclusion of an essay on the College Board English Composition Test, a short-lived experiment by the College Board with a writing sample that students produced under exam conditions and which was forwarded, ungraded, to colleges for evaluation; and the development and dissemination of holistic scoring, a non-red-pen method of global assessment of student writing. Interestingly enough, holistic scoring workshops have generated the same kind of enthusiasm among English teachers as oral proficiency testing workshops have among foreign language teachers.

The model for assessing foreign language proficiency chosen both by the linguists who participated in the ETS Common Yardstick Project and by ACTFL staff, who undertook to carry out the recommendations of the President's Commission, was the ILR Proficiency Definitions, the version of the original FSI scale that had been adopted and subsequently refined over a long period of time by members of government agencies concerned with language training.

From having worked on the ETS Common Yardstick Project, I can speak to the reasons that this particular model was chosen as the one to base a new proficiency assessment scale for academic use. There were two principal reasons:

1. Familiarity. If a "common yardstick" existed already in the US for talking about and measuring second-language proficiency, it was certainly the ILR scale and interview procedure.

2. In the course of the Common Yardstick Project, investigators examined a number of proficiency scales from Australia, Great Britain, and the European continent. There were similarities and differences among them, of course, and what was most striking was that each of them responded to the needs of their particular situation. The investigators were impressed with the 30 years of work on the development, refinement, and use of the ILR scale, and they could find no compelling reason to choose a different scale or to re-invent the wheel by starting a new one of their own.

You are all familiar with the features of the ACTFL/ETS scale, which was the fruit of the Common Yardstick Project, so I will not describe them in detail, but simply list them here for reference:

* Addition of Low and Mid descriptions for Levels 0 and 1 of the ILR scale

* Renaming of the levels -- Novice, Intermediate, Advanced, and Superior, in lieu of 0, 1, 2, and 3-5

* Writing (with the cooperation of ILR member agencies) the Plus and High level descriptions for the first time

- Addition of a proficiency scale for culture
- Creation of language-specific guidelines for French, German, and Spanish
- Deletion in the speaking guidelines of references to listening comprehension

Of these changes, I will comment on here only the expansion of the lower end of the scale with the addition of the Low and Mid descriptions at Novice (0) and Intermediate (1), and the decision to collapse the upper end of the scale, and not distinguish levels of proficiency above Level 3. To begin with the lower end, the Novice and Intermediate Low and Mid descriptions formalized a practice that was already in use at the CIA Language School. Testers officially reported base and plus ratings on its proficiency interviews (0, 0+, 1, 1+, 2, 2+, etc.), but for in-house communication a system of arrows was used to indicate whether an interviewee had given a weak or strong performance within his or her proficiency range. The creation of the Low and Mid verbal descriptions was an attempt to formalize these arrows by attaching verbal descriptions to the knowledge of experienced testers.

The Novice and Intermediate Low and Mid Level descriptions were validated in an informal study conducted by Pardee Lowe at the Language School. From their archives a good number of French and Spanish taped oral proficiency interviews were selected at random -- 15 in each language at the 0 and 1 Levels, for a total of 60 in all. Two groups of experienced testers listened to the tapes in each language: the first group rank ordered the interviews from the strongest performance to the weakest; and the second group compared each of the speech samples to the newly- drafted Novice Low, Novice Mid, Novice High, Intermediate Low, Intermediate Mid, and Intermediate High proficiency level descriptions, to see whether the descriptions provided an adequate way to classify them. When the results of the two groups were compared, it was found that both had classified the speech samples in the same way, thus demonstrating that the Low and Mid descriptions could be used to distinguish among weak and strong performances at the Novice (0) and Intermediate (1) Levels of proficiency.

At the upper end of the scale, the decision was made not to differentiate among performances above Level 3 for the practical reasons: (1) that it was unlikely that academia would produce significant numbers of individuals with proficiency, particularly oral proficiency, above Level 3; and (2) that it was equally unlikely that a significant number of academic oral proficiency testers could be found with the Level 4 and Level 5 oral proficiency needed to test above Level 3. Interestingly enough, DLI was in the midst of deciding at that time to set its own ceiling for oral proficiency testing at Level 4, and not distinguish among Levels 4, 4+, and 5. The rationale, as I understand it, was that instructors did not need proficiency in the language above Level 4 in order to perform their jobs well, and that there was, therefore, no reason to assess their proficiency beyond that point.

2. Experience with the ACTFL Provisional Proficiency Guidelines

The first ACTFL Provisional Proficiency Guidelines were issued at the end of 1982. As they came into use, aspects of them were observed and discussed. First and foremost, there was an irresistible temptation, into which I will no doubt fall in this paper, to focus almost exclusively on the speaking guidelines. This was in part due to the concurrent ACTFL/ETS project to train oral proficiency testers, so that the first exposure that the profession had to the guidelines was for the purpose of assessing the oral skill. It was also due to the fact that the scale for oral proficiency was the best-known and the best-developed within the government, and it was in some sense natural to extend that knowledge and experience to academia. Finally, the speaking guidelines were widely disseminated by the oral proficiency interview test itself.

A second feature of the guidelines that became apparent in the first year or so was that the generic guidelines were not as generic as had been thought originally. As oral proficiency tester training was extended to the less commonly taught languages, especially the non-Western languages, it became apparent that emphases on particular structural features and the relative contribution of the various linguistic factors to each stage of oral proficiency were different across languages in a way not anticipated during the work on the language-specific guidelines in French, German, and Spanish. Perhaps most difficult to grasp initially was the differing weight of sociolinguistic appropriateness, the lack of which makes non-Western language interlocutors squirm a good deal lower on the proficiency scale than is the case for the Western languages.

A third problem with the guidelines is the verbal descriptions themselves. Some problems with wording in the speaking guidelines became apparent during the early tester training workshops; the most striking of these were the descriptions of Novice High, Intermediate High, and Advanced Plus, which communicated a lesser degree of overall performance than was intended. A related, and more serious issue, was that it seemed that the verbal descriptions, no matter how revised, refined, reworded, and annotated, were not complete and clear enough. Indeed, the level descriptions are the imperfect verbal rendering of a vastly complex web of behaviors, competencies, and linguistic features. It is only through administration of the oral interview that testers, who are by and large teachers, not linguists or researchers in second-language acquisition, understand the complexity of the level descriptions. After one has heard 20 or 30 samples of speech at the Intermediate level in Spanish, for example, the seemingly ambiguous phrase from the provisional description, "some evidence of accuracy in basic constructions" is grounded and takes on real meaning. It is not, however, easy to communicate that meaning to others, hence the ongoing quest to refine the wording of the level descriptions to approximate more closely the reality they represent.

Similar kinds of problems have arisen from the language- specific descriptions. The examples in them were meant as examples, not as an exhaustive list, and as estimations of the profiles of language users at a particular level, not as prescriptions of what must be tested or what will be found.

Of the skill areas, the one that loomed largest in terms of conceptual difficulty was culture. As Allen (1985, p. 149) explains it, "the problem does not lie at the practical level of their implementation [i.e. of the cultural guidelines], but at the theoretical level." As Allen (1985, p. 149) sees it, the culture guidelines take as their starting point a conceptual framework of cultural functioning, specifically behavior, without drawing appropriate connections between "behavior in the culture and knowledge about or understanding of the culture." Allen (1985, p. 151) says, in conclusion, that "any conceptual framework developed for culture and for cultural proficiency must take as its point of departure a definition of 'culture' as a field of study, including identification of its component parts and of their relationship." In addition, as for the other skills, "a basis for representing or describing cultural proficiency must be identified, its successive stages established, and a means of evaluation or assessment determined" (p. 151).

The issues that Allen has raised with respect to culture are serious ones, and feature largely in ACTFL's decision to put the culture guidelines on a back burner for the time being. But questions of the same type have been raised with respect to the guidelines in the four skills: what actually is the construct of proficiency? Is there consensus in the profession on what constitutes what Bachman and Savignon (1986) have termed "communicative language proficiency"? Are the criteria given in the ACTFL Proficiency Guidelines an adequate approximation of it? Is the hierarchy of functions, contexts, and elements of accuracy posited by the guidelines the best possible rendering of the progression of global language skill? What is the conceptual framework that binds the level descriptions together, and how can the construct be validated? These are serious questions, ones that proponents, opponents, and skeptics of the so-called "proficiency movement" all agree need to be addressed in the near future.

3. The 1986 ACTFL Proficiency Guidelines

In the years since the provisional guidelines were issued, the academic sector has continued the tradition of our government colleagues by adjusting and refining the verbal descriptions themselves. The 1986 guidelines represent the most concentrated effort to do so, and were undertaken to accomplish the following:

(1) to address and correct the derivative nature of the guidelines for the listening, reading and writing skills;

(2) to correct the Western-language bias in the English- language guidelines, both to make them more truly "generic" and to enable writers of language-specific guidelines to use them as a basis for their work;

(3) to adjust the wording of the guidelines to make more transparent the concept of language proficiency that underlies the descriptions; and

(4) to correct infelicities in wording, such as the Novice High, Intermediate High, and Advanced Plus descriptions for speaking, which gave an impression of lesser proficiency than was intended.

A comparative reading of the provisional guidelines and the revised 1986 guidelines reveals the following features of the more recent version:

(1) Each level is introduced by an overall description, which provides a framework for the elaboration of the sub-ranges within each level, as well as communicates the essential features of performance at each level. The goal here is to make the wording of the level descriptions a closer approximation of the reality in the minds of the writers.

(2) A "Distinguished" level of proficiency was added for the receptive skills. This was done in recognition of the fact that learners progress more rapidly in the receptive skills than in the productive skills, and that considerable numbers of students, particularly at the graduate level, surpass the minimum requirements for the Superior level in listening and reading skills. The Distinguished rating refers to a level of proficiency that is both linguistic and cultural, such that the listener or reader can comprehend texts "from within," in the words of the guidelines, the cultural framework in which they were produced.

(3) The culture guidelines themselves were deleted, due in large part to an acceptance of the case made by Allen and others that a better-defined conceptual framework must precede the elaboration of proficiency guidelines. As things currently stand, there is still no consensus in the profession about the areas of knowledge and understanding that should be included; in addition, a convincing case for a hierarchy of knowledge and skills in the area of culture still remains to be made.

(4) All of the descriptions have been reworded to put greater emphasis on what language users can do, in contrast to what they cannot do. The descriptions still identify

how language users fall short, and why they should be rated at a particular level and not the next higher one, but the verbal emphasis has shifted. This has particular relevance to those who are interested in using the guidelines as a point of departure for setting curricular goals.

(5) Some of the thornier wording problems have been corrected. Consider, for example the description for Intermediate Mid speaking proficiency:

> Some evidence of grammatical accuracy in basic constructions, for example, subject-verb agreement, noun adjective agreement, some notions of inflection.

Compare it to the description of the same level in the revised 1986 guidelines:

> Speech may continue to be characterized by frequent long pauses, since the smooth incorporation of even basic conversational strategies is often hindered as the speaker struggles to create appropriate language forms.

The provisional description contains the problematic reference to "some evidence of grammatical accuracy," which had been found to cause confusion at testers' workshops, and gave examples of "basic constructions" that were applicable only to some languages. The revised description makes reference to "appropriate language forms" rather than mention particular forms or structures, thus becoming relevant over a larger number of languages. Similarly, the provisional form of the Advanced level description for listening comprehension features numerous examples of text topics that the listener might be presented with, specifically targets face-to-face conversation as the context for listening comprehension, and contains wording that is unfortunately derivative from the speaking guidelines: " . . . delivered at a normal rate with some rewording by a native speaker not used to dealing with foreigners."

The 1986 version, on the other hand, corrects these infelicities. The description of Advanced level listening comprehension features prominently the role of "topic familiarity" in comprehension, and mentions types of texts that the listener might be exposed to, such as interviews, short lectures, and news items, without specifying the possible content of such texts. The strong implication from the text types mentioned is that one-way listening is now included, in addition to conversation.

In spite of these major improvements in the guidelines, some problems still remain. I will mention two of them: First, despite the rewording, the level descriptions are still not transparent enough, and are still causing misunderstanding in readers who may not have "lived and breathed" them in the way that oral proficiency testers have experienced the speaking guidelines. For example, Bachman and Savignon (1986, p. 386) reproduce the ACTFL Advanced level speaking description and the ILR Level 2 description, discuss the differences in the wording, and express the concern that these differences will "create a lack of comparability of ratings, since it is not certain that interviewers using the different scales will try to elicit the same functions."

The concern for comparability expressed by Bachman and Savignon is certainly an important one, but the differences in wording between the ACTFL and ILR level descriptions for speaking do not represent differences in functional ability, and would not induce trained testers to elicit different functions. The problem, addressed indirectly here, is a practical one: it is time-consuming and often expensive to become an accurate and elegant oral proficiency interviewer and rater. To my knowledge, no one has learned how to do oral proficiency interviews across the whole proficiency scale simply by reading and understanding the descriptions of the levels; the words alone do not, and perhaps cannot communicate the reality of communicative proficiency. There seems to be no shortcut around intensive training workshops that expose the participants to numerous speech samples and engage them in extensive discussion of the relationship between those speech samples and the words of the level descriptions.

Bachman and Savignon's concerns, however, do speak to the need to carry out comparability studies between ILR and academia, for both interviewers and raters. Much of the concern in the area of comparability has been with inter-rater reliability and, to a lesser degree, with intra-rater reliability over time. I would like to add to the agenda the need for inter-interviewer reliability studies, to investigate whether trained testers do indeed elicit comparable speech samples from the same group of speakers.

A second problem that still remains in the guidelines, a problem that we have inherited from our ILR colleagues, is their bias in favor of professional, white-collar diplomatic language. That orientation is certainly appropriate for the diplomatic context for which the proficiency scale and interview procedure were originally created, a context whose ultimate goal was to prepare individuals to operate successfully in international negotiations.

In this regard, I agree with Bachman and Savignon that the level descriptions do not constitute as universal a definition of proficiency as we might like. In ETS' involvement with the Peace Corps, for example, we have seen over the years that white-collar professional language is often not an appropriate goal for volunteers to aspire to, and in fact might well be a detriment to carrying out one's job in the field.

The Distinguished level of proficiency for the receptive skills in the 1986 guidelines does mention literary texts and address some of the uses of language that are important to academia, but that are absent in the ILR scales, such as awareness of the aesthetic properties of language, sensitivity to affective overtones, and appreciation of stylistic nuance and subtlety. Much work is yet to be done to develop criteria for communicative proficiency that belong to no particular walk of life or, if that proves unfeasible, to decide which aspects of communicative proficiency are appropriate for various user purposes.

4. Implications of the Guidelines

The guidelines were originally developed to assess learners' performance, not to instruct them. But the backwash effect of testing on the instructional process is a well-known phenomenon, and finds no exception here. In the government language schools, curricula, materials, and teaching strategies are organized so that students move more quickly and efficiently up the proficiency scale. This movement, if not fully equated with language learning itself, is certainly viewed as a close approximation thereof.

In academic settings, the guidelines and, especially, the oral proficiency interview, have begun to have a backwash effect on instruction which, in my opinion, has been most positive. I certainly agree that the last word on proficiency and on the hierarchy of functional tasks posited by the ACTFL Guidelines is not in, but in the meantime the effect of proficiency and the proficiency guidelines on teaching practice has been beneficial. Indeed, proficiency tests, which by their summative nature and, in the case of the productive skills, labor-intensive format, are given seldom in a student's language-learning career, may not ultimately have their greatest impact as evaluation procedures, but rather, in Bragger's (1986, p. 11) words, as "agents of change."

In terms of curriculum planning, the ACTFL Proficiency Guidelines are not a curriculum guide, or anything close to it. They represent a way of visualizing curriculum as a three-dimensional, expanding spiral, rather than a linear progression toward increased ability to function in the target language. The proficiency-oriented curriculum is seen as constantly re-incorporating material from earlier stages in different contexts, for different purposes, so that students' abilities and knowledge are constantly expanding from an already-familiar, already-mastered base.

This vision of vertical integration via a spiraled curriculum is, of course, not new. But what the proficiency guidelines add is a new perspective, one that teachers, particularly at the secondary level, have perhaps been more able to grasp. I believe that it is the oral interview, in its ability to make the concept of proficiency more tangible, that has enabled language practitioners to understand the rhythmic nature of a spiraled curriculum and begin to take steps to put it into practice.

Related to this idea of vertical integration is the notion of horizontal integration among the three facets of proficiency --function, context, and accuracy. The framework of proficiency demands that course planning pay as much attention to the why, the where, and the when of language use as to the how well. As we all know, this is easier to accept in principle than in reality; paradigms of all kinds are easier to divide up into class-length segments, easier to explain, easier to drill, and easier to test than to integrate with culturally authentic linguistic tasks in realistic contexts.

Once again, this is not to say that the proficiency guidelines alone invite us to integrate language with culture, or accuracy with functions and contexts, but rather that the guidelines may be one more in a series of voices inviting us to do so.

If we consider now what the proficiency orientation may have to offer at the level of day-to-day teaching strategies and classroom testing, its greatest contribution may well be balance --balance between skill-getting and skill-using activities, between the four skills and culture, between teacher-centered and student-centered approaches, between formative and summative evaluation (Bragger, 1986).

One of the potentially most valuable implications of the proficiency guidelines for classroom practice is in the area of achievement testing. Under the influence of national standardized tests and the grammar-heavy chapter and unit tests supplied by textbook publishers, many of us, especially those of us who teach large numbers of students in language courses, have fallen into the habit of having students translate sentences, fill in blanks, and define vocabulary words. The proficiency guidelines, particularly their insistence on the coexistence of function, context, and accuracy, provide us with ideas to broaden the scope of our achievement tests, making them more integrative while at the same time still measuring students' short-term mastery of course material.

5. Conclusion

In conclusion, I have attempted to address in this paper the pedagogical backdrop for the creation of the ACTFL Provisional Guidelines, to examine their strengths and weaknesses, and to discuss their implications for curriculum development and classroom testing. The research agenda proposed by Clark and Clifford (1987) should receive a high priority as we move forward to establish and validate a construct of communicative language proficiency for both the academic and government sectors. In the meantime, my strong recommendation is that we proceed as well to support classroom practices based on the concept of proficiency, practices that enhance students' communicative skill and enthusiasm for language study.

References

Allen, W. (1985). Toward cultural proficiency. In A. Omaggio (Ed.), *Proficiency, curriculum, articulation: The ties that bind* (pp. 137-166). Middlebury, VT: Northeast Conference.

Bachman, L., & Savignon, S. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. *Modern Language Journal, 70*, 380-390.

Bragger, J. (1986). Teaching for proficiency: Are we ready? *ADFL Bulletin 18*, 11-14.

Clark, J., & Clifford, R. (1987). The FSI/ILR/ACTFL proficiency scales and testing techniques: Development, current status, and needed research. This volume.

Hiple, D. (1986). A progress report on the ACTFL proficiency guidelines, 1982-1986. In H. Byrnes and M. Canale (Eds.), *Defining and developing proficiency: Guidelines, implementation, concepts* (pp. 5-24). Lincolnwood, IL: National Textbook Company.

PROBLEMS IN EXAMINING THE VALIDITY OF THE ACTFL ORAL PROFICIENCY INTERVIEW

Lyle F. Bachman
University of Illinois at Urbana-Champaign

The primary problems in measuring speaking ability through an oral interview procedure are not those related to efficiency or reliability, but rather those associated with examining the validity both of interview ratings as measures of ability in speaking and of the uses that are made of such ratings. In order to examine all aspects of validity, the abilities measured must be clearly distinguished from the elicitation procedures, in both the design of the interview and in the interpretation of ratings.

Research from applied linguistics and language testing is consistent with the position that language proficiency consists of several distinct but related abilities. Research from language testing also indicates that the methods used to measure language ability have an important effect on test performance. Two frameworks--of communicative language ability and of test method factors--are proposed as a basis for distinguishing abilities from elicitation procedures, and for informing a program of empirical research and development.

The validity of the ACTFL Oral Proficiency Interview as currently designed and used cannot be adequately examined, much less demonstrated, because it confounds abilities with elicitation procedures in its design, and it provides only a single rating, which has no basis in either theory or research. As test developer, ACTFL must accept responsibility providing sufficient evidence of validity to support uses of ratings, in anticipation of possible litigation.

1. Introduction

The ability to speak a foreign language is arguably both the most important and the most problematic of the abilities we attempt to measure. Speaking is the most salient of the "four skills" in the sense that whatever the emphasis of the particular method, there is an intuitive "lay" perception among learners that being able to speak the language is the primary, or central goal in language learning.

Measuring ability in speaking is problematic because of the complexity of both the skills involved and the context in which these skills are to be elicited and assessed. That is, the major problems in measuring speaking ability are *not* those related to efficiency, such as the time and effort required to elicit a speech sample and score it. As we all know, it is quite possible for a classroom teacher to efficiently elicit and reliably rate the speech samples of his students. Indeed, as has been demonstrated over the years in a wide range of contexts (e.g., Adams, 1978; Lowe, 1978; Mullen, 1978, 1980; Clifford, 1978; Clark, 1978), and more recently with the ACTFL Oral Proficiency Interview (OPI), a high degree of efficiency and reliability can also be obtained with multiple examiners and with large groups of subjects (Magnan, 1986).

The problems to which I refer, and to which this paper is addressed, are those associated with demonstrating evidence for the validity of interview ratings. I will first argue that in order to provide a basis for making valid inferences about ability in speaking, any measure of this ability must clearly distinguish the ability to be measured from the methods or procedures used to elicit evidence of this ability. In order to examine *content validity*, traits and methods must be distinct in the test design; in order to examine *construct validity*, they must also be distinguishable in the performance of test takers. Next, I will briefly describe two frameworks that I believe will be useful both for characterizing abilities and test methods, and for distinguishing between them. I will then argue that the validity of OPI ratings as indicators of communicative language ability cannot be adequately examined, much less demonstrated. Finally, I will make some suggestions for redesigning the OPI, and urge both ACTFL and users of the OPI to initiate a program of research aimed at investigating both the validity of the ratings and the uses for which they are appropriate.

2. Distinguishing abilities from test methods

Test design and content validity – Content validity concerns "the degree to which the sample of items, tasks, or questions on a test are representative of some defined universe or domain of content" (APA, 1985, p. 10). It is generally understood that in order to examine the content validity of a test, the *ability* domain must be specified clearly, completely and unambiguously. What is not so widely recognized is that the measurement procedure must also be described "with such clarity that other investigators could reproduce the significant aspects of the procedure for themselves" (Cronbach, 1971, p. 449). Cronbach (1984) discusses several aspects of the measurement procedure and indicates that these must all be specified in order for the content validity of the test to be examined. These aspects include the range of tasks, stimuli or situations included in the test, the instructions that describe the test taker's task, the type of response expected of the test taker and how these responses are to be scored (Cronbach, 1984, p. 145).

The importance of specifying both the content or ability domain and aspects of the measurement procedure in the test design is also recognized in the literature on criterion-referenced test development. Popham, for example, includes the following as components to be specified in test design: 1) "what it is that the test measures," 2) the attributes of the stimuli which will be presented to the test taker and 3) the nature of the responses the test taker is expected to make (Popham, 1978, pp. 121-122). Hambleton (1984), likewise, in his discussion of validity of criterion- referenced tests, relates test design considerations directly to content validity.

In examining content validity, then, we are concerned not only with the domain of abilities measured, but with the measurement procedures as well. This is clear from Cronbach's description of validation:

> a validation study examines the procedure as a whole. Every aspect of the setting in which the test is given and every detail of the procedure may have an influence on performance and hence on what is measured. Are the examiner's sex, status, and ethnic group the same as those of the examinee? Does he put the examinee at ease? Does he suggest that the test will affect the examinee's future, or does he explain that he is merely checking out the effectiveness of the instructional method? Changes in procedures such as these lead to substantial changes in ability- and personality-test performance, and hence in the appropriate interpretation of scores. (1971, p. 449)

Performance on tests and construct validity – Construct validity concerns the extent to which performance on tests is consistent with predictions that we make on the basis of a theory of abilities, or constructs. We would predict, for example, that individuals who have high levels of ability in grammar and low ability in, say, reading, would perform well on different tests of grammar, such as a multiple-choice test or a rating of grammatical accuracy on an oral interview, and would receive low scores on different tests of reading. If we found that individuals performed as expected on these tests, this would be evidence for their construct validity.

Construct validity is often mistakenly considered to be of importance only in situations where content validity cannot be examined because the domain of abilities is not well-specified. With reference to language tests, this misconception has generally been applied to the distinction between measures of proficiency and of achievement. The problem with this view, however, is that it ignores the major limitation of content validity. That is, demonstrating that the contents of a test accurately represent a given domain of ability fails to take into consideration, or to account for, how individuals actually perform on the test (Cronbach, 1971; Guion, 1977; Fitzpatrick, 1983). Thus, even though the content of a given test does not vary across different groups of individuals to which it is given, the performance of these individuals is likely to vary considerably. Furthermore, the score of an individual on a given test is not likely to be equally valid for different uses. For example, although a rating based on samples of oral language use may be valid as an indicator of speaking ability, it may not be valid as a measure of teaching ability. The primary limitation of content validity, then, is that it focuses on tests, rather than test scores. This limitation has been delightfully characterized by Messick in an extended simile:

> content validity is like the barker outside a circus tent touting two bowing aerialists, a waving clown, and a poster of a lady riding a unicorn as a sample of the show you will see inside. It is not just that the sample is not representative of the variety of

> individuals and animals inside or even an accurate portrayal of them, it is that you do not see any performances. (1975, p. 961)

Many psychometricians (e.g. Cronbach, 1971, 1980, 1984, 1985; Loevinger, 1957; Messick, 1975, 1981a; Linn, 1980) have argued that construct validation is essential in any situation in which we wish to make inferences about an individual's ability on the basis of test scores, and the centrality of construct validation has been codified by the measurement profession as a whole in the Standards for Educational and Psychological Testing (APA, 1985). While a number of approaches may be followed in investigating construct validity, all examine the extent to which test scores demonstrate convergence and discrimination. Convergence is demonstrated when the measure is "substantially related to other measures of the same construct and to other variables that it should relate to on theoretical grounds" (Messick, 1975, p. 956). Discrimination, on the other hand, is demonstrated when the measure in question is not substantially related to measures of other distinct constructs. The classic statement of the requirements for empirically examining construct validity was provided by Campbell and Fiske (1959):

> Each test or task employed for measurement purposes is a *trait-method* unit, a union of a particular trait content with measurement procedures not specific to that content. The systematic variance among test scores can be due to responses to the measurement features as well as responses to the trait content. . . . Validity is represented by the agreement between two attempts to measure the same trait through maximally different methods. (pp. 81, 83)

From this statement it is clear that in order to examine construct validity we must not only be able to specify what are similar and dissimilar traits and what constitute maximally distinct methods, but we must also be able to distinguish traits from methods.

The effects of test method – In obtaining a rating of oral language ability, we elicit a sample of an individual's performance *under certain conditions*. These conditions, or test method factors, are distinct from the language abilities or competencies we wish to measure. There is considerable evidence that such method factors have a significant effect on performance on language tests, and hence, upon test scores. Clifford (1978), for example, found very high correlations among oral interview ratings of different traits, indicating a strong effect of test method on performance. Test method factors are thus potential sources of error for every rating we obtain, and consequently limit both their reliability and validity as indicators of the ability we want to measure.

The effects of test method present two problems: 1) a dilemma in choosing the type of error we want to minimize, and 2) ambiguity in the inferences we can make from ratings. If we permit method factors to vary from interview to interview in an uncontrolled manner, they are sources of random measurement error. One way to minimize random error is to standardize the procedures, thereby controlling the conditions under which the interview is given so that they are the same for all interviews. In the OPI, this standardization involves carefully training interviewers so that they: 1) follow the same elicitation procedures; 2) use the same definitions for rating the performances of different individuals; and 3) administer the interview in the same environment and in approximately the same amount of time. However, the problem created by standardization is that controlling these method factors causes them to have a systematic effect on test scores. That is, they affect the performance of all individuals who are interviewed, and thus constitute potential sources of test bias.[1] The dilemma, then, is that by standardizing test method factors we reduce *random* errors, thereby increasing reliability, but in so doing, we introduce *systematic* errors, thereby decreasing validity. If this systematic effect of test method is sizeable, therefore, it clearly limits the validity of the rating as an indicator of the individual's language ability.

As Stevenson (1981) has noted, the interpretation of language test scores is problematic, since traits are frequently difficult to distinguish from methods. This is particularly true with "performance" tests, such as oral interviews, in which the modality (productive) and channel (oral/aural) of the ability (speaking) match the modality and channel of the elicitation procedure, or test method, thereby making it difficult to clearly distinguish ability from test method.[2] Thus it is not clear whether ratings of grammar and sensitivity to register, for example, can be interpreted as indicators of these language abilities, or as indicators of an individual's ability to perform well under a particular set of test method conditions. In two studies, Bachman and Palmer attempted to answer this question. In one study (Bachman & Palmer, 1981), they found sizeable loadings on factors associated with test method oral interviews, while in the other (Bachman & Palmer, 1982) they found that factor loadings for the interview method were consistently higher than factor loadings associated with the traits being measured.

This research suggests two options for interpreting scores from oral interviews. First, we can view the effect of test method as a source of error that reduces both the reliability and validity of ratings. This interpretation has been the most common among language testing researchers to date. The alternative is to recognize the capability of oral interview ratings for providing reliable indicators of more specific skills. That is, rather than viewing a rating of, say, grammatical competence based on an oral interview as a relatively unreliable indicator of this broad ability, we might interpret it as a highly reliable indicator of an individual's skill in using grammatical structures accurately *in the contexts and under the conditions that are included in the testing procedure.* In this regard, I would recall Higgs and Clifford's (1982) injunction that when we make global judgments of proficiency "we are obligated to reveal to our students the limitations of such judgments, lest we deceive them and ourselves. We must tell them not that they are competent to speak German, but that they are competent to meet routine physical and social obligations in an environment where German is spoken" (Higgs & Clifford, 1982, pp. 60-61).

3. Factors that affect performance on language tests

From the above discussion, it is clear that any framework for describing factors that affect performance on language tests must specify both the domain of language abilities and the test method characteristics that may systematically affect performance on language tests.

Language proficiency: Unitary or divisible? – The nature of the ability we wish to measure is a fundamental consideration in the development and use of all tests, and it is therefore not surprising that language testing researchers over the years have sometimes appeared to be preoccupied with problems associated with empirically validating this or that theory of language ability. Although empirical research examining the nature of language abilities dates at least from the 1940's with John B. Carroll's pioneering work (Carroll, 1941), it was John Oller's "unitary trait hypothesis" and his "g-factor" studies that focussed the attention of language testing researchers on this question and brought it to the attention of language teachers (Oller, 1976, 1979). According to the unitary trait hypothesis, language proficiency consists of a single, global ability. The primary evidence provided in support of this hypothesis consisted of factor analytic studies which found that scores on a wide range of language tests consistently loaded most heavily on a single factor that came to be called the "g-factor", for "general language proficiency."

Although Oller's work had an immediate impact on language testing research, and is still cited in some language teaching circles, it was soon recognized that the unitary trait hypothesis was untenable: the g-factor studies were flawed methodologically (Carroll, 1983; Vollmer & Sang, 1983), and subsequent studies (e.g., Bachman & Palmer, 1981, 1982; Upshur & Homburg, 1983) provided evidence invalidating this hypothesis. After considering the evidence, Oller, with admirable scholarly objectivity and honesty, stated in 1983 that "the unitary hypothesis was wrong." (p. 352).

The evidence from language testing research is generally consistent with the hypothesis that language proficiency consists of several distinct abilities that are either related to each other or which are related to a higher order, general ability (Carroll, 1983; Oller, 1983). Many language testing researchers are thus focusing their efforts on identifying the component abilities of language proficiency and are now for the most part working within an expanded framework of "communicative language ability" of which the major distinguishing characteristic is its recognition of the importance of context beyond the sentence level to the appropriate use of language. This context includes both the discourse of which individual sentences are part and the sociolinguistic situation which governs, to a large extent, the nature of that discourse, in both form and function.

Along with this recognition of the context in which language use takes place has come a recognition of the dynamic interaction between that context and discourse, and an expanded view of communication as something more than the simple transfer of information (e.g., Johnson, 1982; Savignon, 1983). Frameworks of communicative language ability thus provide a much more inclusive description of the abilities required to use language effectively than do earlier models, in that they include, *in addition* to the knowledge of grammatical rules, the knowledge of language functions, or illocutionary acts and of sociolinguistic conventions, and the recognition of language use as a dynamic process.

Communicative language ability – Incorporating earlier work on communicative competence (e.g., Hymes, 1972, 1982; Savignon, 1972, 1983; Canale & Swain, 1980; Widdowson, 1978; Johnson, 1982), and on communication strategies (Faerch & Kasper, 1983), I propose to define communicative language ability (CLA) as comprising both knowledge, or competence, and skill in implementing, or executing that competence (Bachman, forthcoming). CLA consists of three components: language competence, strategic competence and the psychophysiological mechanisms required to implement these competencies in language use. This framework is illustrated in Figure 1.

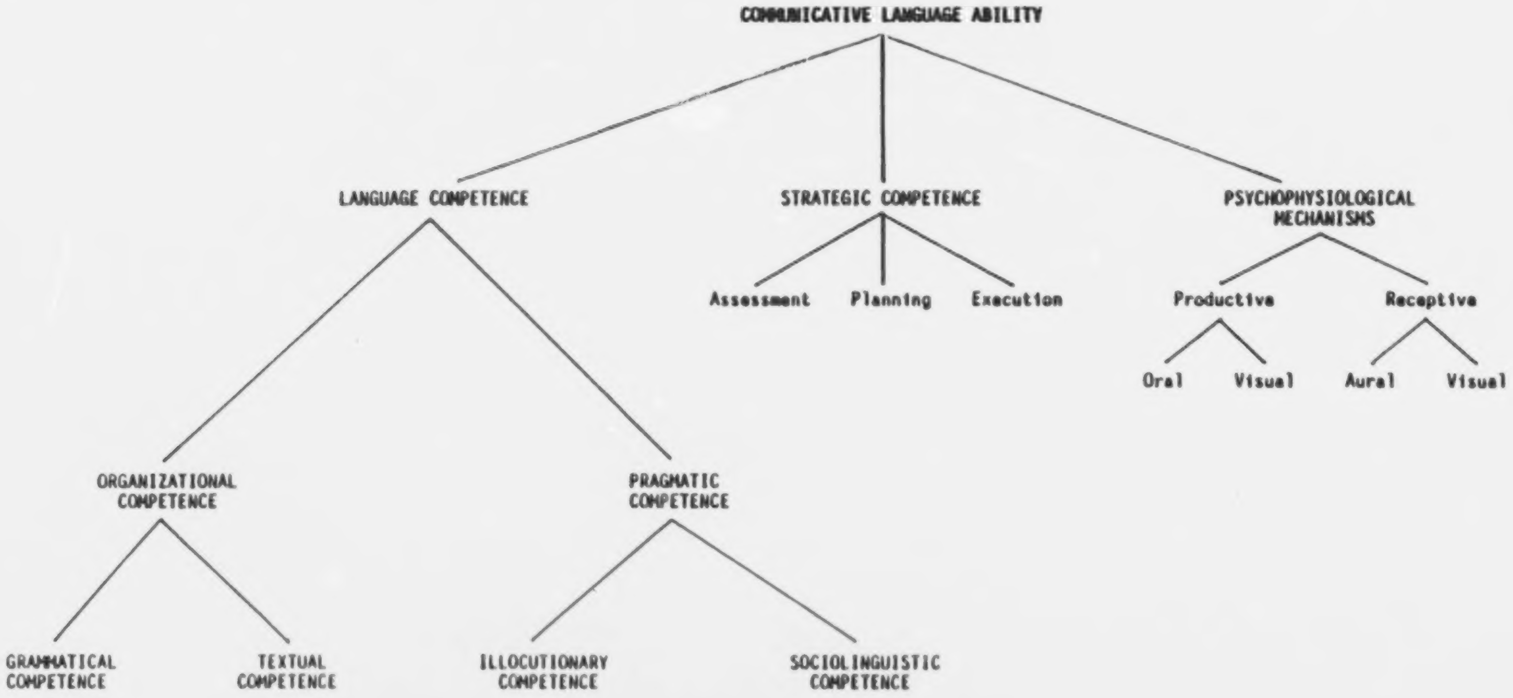

COMMUNICATIVE LANGUAGE ABILITY
LANGUAGE COMPETENCE
STRATEGIC COMPETENCE
PSYCHOPHYSIOLOGICAL MECHANISMS
Assessment
Planning
Execution
Productive
Receptive
Oral
Visual
Aural
Visual
ORGANIZATIONAL COMPETENCE
PRAGMATIC COMPETENCE
GRAMMATICAL COMPETENCE
TEXTUAL COMPETENCE
ILLOCUTIONARY COMPETENCE
SOCIOLINGUISTIC COMPETENCE

Figure 1

A Framework for Describing Communicative Ability

Language Competence

Language competence can be classified into two types: organizational competence and pragmatic competence. *Organizational competence* comprises those abilities involved in controlling the formal organization of language for creating or recognizing grammatically correct sentences, comprehending their propositional content, and ordering them to form texts. These abilities are of two types: grammatical and textual. *Grammatical competence* includes rules of lexis, morphology, and syntax, which govern the choice of words to express specific significations, their forms, and their arrangement in sentences to express propositions. *Textual competence* includes the knowledge of the conventions of cohesion and rhetorical organization for joining utterances together to form a text.

Pragmatic competence includes those abilities which, in addition to organizational competence, are employed in the contextualized performance and interpretation of socially appropriate illocutionary acts in discourse. Pragmatic competence thus includes *illocutionary competence*, or the knowledge of how to perform illocutionary acts, or language functions, and *sociolinguistic competence*, or knowledge of the sociolinguistic conventions which govern appropriate language use in a particular culture and in varying situations in that culture.

Psychophysiological Mechanisms

Communicative language ability may be realized in listening, speaking, reading and writing. These skills can be categorized in terms of *modality* (receptive, productive) and *channel* (auditory, visual), and distinguished by the psychophysiological mechanisms that are involved in language use. Thus in the receptive modality (listening, reading) sensory mechanisms are employed, while in the productive modality (speaking, writing), motoric mechanisms are employed.

Strategic Competence

Communicative language use involves a dynamic interchange between the language user, the discourse, and the context of the situation in which the use occurs. Interpreting (producing or understanding) discourse thus requires the ability to assess the context for information relevant to the communicative goal and to then match information in the discourse to this information. The matching of new information with relevant information that is available in the context (including presupposition and real world knowledge) and in the discourse itself, and mapping this onto the maximally efficient use of existing language abilities is a function of strategic competence.

In their discussion of communication strategies, Faerch and Kasper (1983) present a model of speech production that includes a communicative goal, a planning process, a plan, and the execution of that plan through neurological and physiological processes. While Faerch and Kasper's model is intended to explain the use of communication strategies in interlanguage communication, I believe it provides a basis for a more general description of strategic competence in communicative language use. I would define strategic competence as the ability that enables us to 1) *formulate a plan* for realizing a particular communicative goal, 2) *execute that plan* as an utterance, and 3) *assess* the extent to which the communicative goal has been achieved. These functions are represented schematically in Figure 2.

Test method factors – The most detailed treatment of test method factors in the literature on language testing to date is Carroll's 1968 paper, in which he discusses four general types of language test tasks, which are distinguished largely in terms of

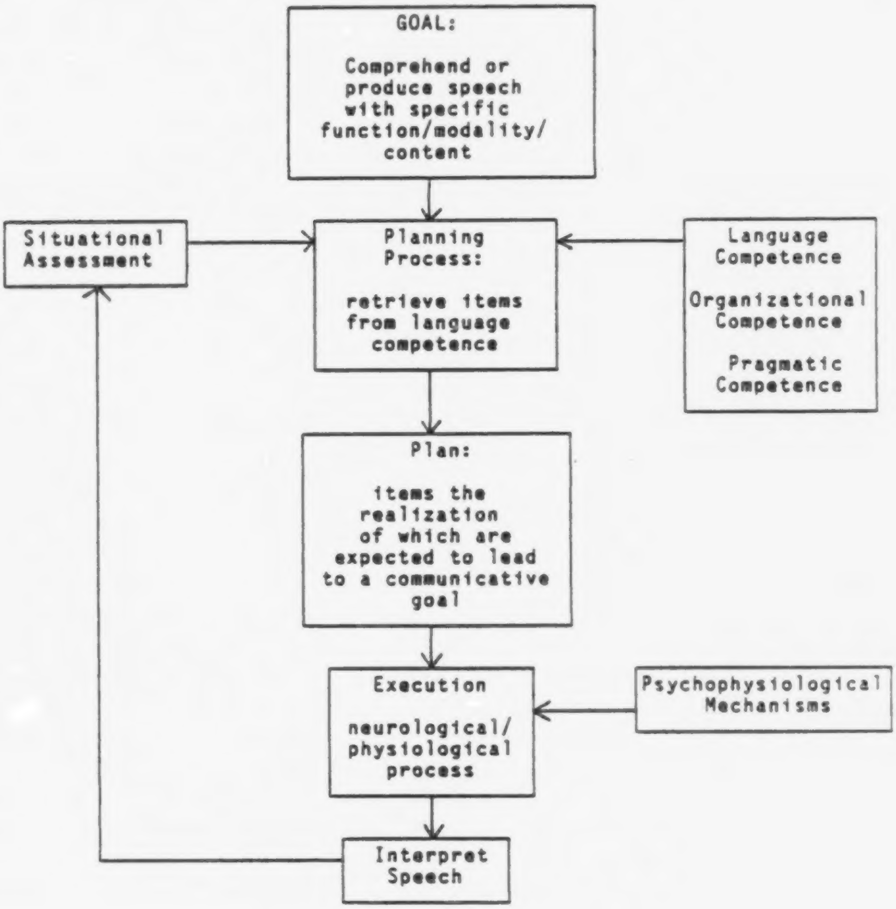

GOAL:
Comprehend or produce speech with specific function/modality/ content
Situational Assessment
Planning Process:
retrieve items from language competence
Language Competence
Organizational Competence
Pragmatic Competence
Plan:
items the realization of which are expected to lead to a communicative goal
Execution
neurological/ physiological process
Psychophysiological Mechanisms
Interpret Speech

Figure 2

A Model of Language Use

(Adapted from Faerch & Kasper 1983)

differences in their stimulus and response characteristics. I would propose a framework for characterizing test method factors that extends Carroll's taxonomy, incorporating more recent views of the nature of language (Bachman forthcoming). This framework includes the following categories: 1) the testing environment, 2) the nature of the test instructions, 3) the nature of the input the test taker receives, 4) the nature of the response to that input, and 5) the interaction between input and response.

The *testing environment* includes factors such as the time of testing and the familiarity of the place, equipment and the personnel administering the test. *Test instructions* include factors such as the the specification of the task and the explicitness of the criteria for correctness. The *input* includes the following factors: 1) channel (auditory, visual), 2) format (type of presentation, type of language sample, specification of problem, speededness), and 3) the nature of the discourse of the stimulus (propositional content, textual characteristics, illocutionary characteristics). The *response* includes the following factors: 1) channel, 2) type of response (identification, selection of correct alternative, production), 3) the form of the response (language, nonlanguage), 4) the nature of the discourse of the response, and 5) restrictions on response, such as requiring the use of specific grammatical forms or organizational patterns in a test of writing.

These frameworks of communicative language ability and test method factors are not presented as a complete theory of language test performance, but are rather intended to stimulate and inform research and development. I believe they are specific enough to guide scale development and to generate hypotheses for validation studies. As such, they are themselves subject to validation, and will in all likelihood change to reflect our growing knowledge.

4. Summary

The thrust of this paper has been to examine the extent to which the ACTFL Oral Proficiency Interview, as embodied in the ACTFL Proficiency Guidelines (ACTFL, 1986), provides ratings that are valid indicators of the ability to communicate through speaking. In my opinion, the validity of interpretations and uses of scores from the OPI has not been demonstrated, nor can it be as the scale is now designed. Two characteristics of the current procedure make validation a virtual impossibility. First, as Bachman and Savignon (1986) argue, traits and test methods are confounded in both the design of the interview and in the interpretation of ratings. While proponents of the procedure (e.g., Higgs & Clifford, 1982; Liskin-Gasparro, 1984; Lowe & Liskin-Gasparro, n.d.) argue that this is a major strength of the OPI, I view this confounding of function, content and structure as a fundamental flaw that makes validation impossible. A second weakness, in my opinion, is the use of a single, global rating of oral language proficiency that has no basis in either current linguistic theory or empirical research in language testing. The inadequacy of a single score for examining the validity of a measure as an indicator of ability, as opposed to its capacity for predicting some future performance, has also been argued by Upshur (1979).

The fundamental requirement for validation is that we must begin with definitions of language ability and testing procedures that provide both direction for test design and hypotheses about test performance. I have presented frameworks that identify specific components of communicative language ability and test methods, and believe these are consistent with current developments in linguistic theory, language teaching and language testing. And while these particular frameworks should be viewed as "provisional", they do, I believe, provide an adequate basis for designing a program of validation research.

5. Suggestions for research and development on the ACTFL Oral Proficiency Interview

Specific suggestions for research and development on the OPI have been made elsewhere (Bachman & Savignon, 1986; Bachman & Clark, 1986), and I will not repeat all of these here. Of critical importance to examining the validity of the OPI, however, are the following:

1) Abilities must be clearly distinguished from test methods – I believe the best way to accomplish this is to define the rating scales abstractly, in terms of the degree of mastery of the various abilities themselves, without reference to specific functions, structures or content domains. This implies the development of separate scales for the components of communicative language ability that are to be measured.

In addition to defining the domain of abilities, the domain of test method factors must be clearly specified. This apparently has been done, to a large extent, with elicitation procedures, through the program of training associated with the OPI. What remains to be done, I believe, is to more clearly specify the contexts and content areas that are part of the elicitation procedure. In this regard, I believe the work of Childs (1987) on types of texts and the program of research described by Dandonoli (1987) with respect to listening and reading are steps in the right direction.

2) ACTFL must make a commitment to a program of research into the validity and appropriate use of the oral interview – In preparing the Guidelines, training interviewers, and disseminating the testing procedure to the profession, ACTFL has taken the role of test developer. However, it has not as yet taken seriously the developer's responsibility for demonstrating the validity of the interpretations and uses that are made of OPI ratings. This responsibility is clearly stated in the Standards for Educational and Psychological Testing (APA, 1985), which are accepted not only as guiding principles by the measurement profession as a whole, but have also provided a basis for legal arguments regarding the uses of tests.

This program of on-going research must address both the meanings and values attached to test scores (Messick, 1975). That is, research must be directed not only to demonstrating the validity of OPI ratings as indicators of communicative language ability, but also to investigating the uses for which these ratings are valid (Messick, 1981b). If ACTFL itself does not have the resources to conduct this program of research, it should take an active role in identifying competent measurement specialists in the language teaching profession, involving them in the design and conduct of appropriate research and development activities, and seeking sources of funding to support such activities.[3]

6. Conclusion

Are ratings from the OPI valid indicators of communicative language ability? As the OPI is currently designed and used, I believe they are not. This is not to say, however, that they cannot be. As its proponents correctly observe, the OPI is based on procedures that have developed over years of practice, and about which a great deal of experience has accumulated. However, this experience in no way constitutes adequate evidence for validity. What the long experience of government agencies with the OPI does provide is the following: 1) evidence that interview ratings give some prediction about an individual's future performance in positions with the government, and 2) circumstantial evidence, if you will, that the ratings measure some ability that is of interest to the government's language schools and, hopefully, to their students and clients. Whether this type of predictability is relevant, and whether this ability, whatever it is, is of interest to the wider needs of the academic language teaching community are questions that can no longer be begged by reference to experience.

The "Proficiency Movement" has infused new interest and enthusiasm into foreign language teaching, and this is to be commended. However, the success of the OPI is a two-edged sword. In an increasingly litigious society, both ACTFL and those who use the OPI must be prepared to demonstrate evidence that a rating is both a valid indicator of ability in speaking, and valid for the purpose of making decisions about, say, minimum competency or employability. Thus, as ratings from the OPI come to be increasingly used to make decisions about individuals--both language learners and language teachers--the question the foreign language teaching profession must address is, "What will be our defense when we are held accountable for those decisions?" The complexities of addressing this question have been discussed by Cronbach (1980) and Messick (1981b), among others. The consequences of ignoring it can be found in the annals of our judiciary system.[4]

Notes

[1] This systematic effect may be general, in that it affects all individuals equally, or specific, in that it affects some individuals more than others. (Kane, 1982)

[2] The failure to distinguish traits from methods was characteristic of much of the research that led to the formulation of the "unitary trait hypothesis" of language proficiency (Oller, 1979).

[3] It is encouraging to see that ACTFL is beginning to understand the need for research, if the program described by Dandonoli (1987) is indeed a reality.

[4] Of particular interest with respect to the possible use of the ACTFL oral interview for minimum competency testing is the landmark decision handed down in the Debra P. V. Turlington case. The implications of this decision, for both the measurement and the educational concerns of testing, are discussed at length in the papers included in Madaus (1983).

References

American Council on the Teaching of Foreign Languages. (1986). *ACTFL Proficiency Guidelines*. New York: Author.

Adams, M. L. (1978). Measuring foreign language speaking proficiency: A study of agreement among raters. In J. L. D. Clark (Ed.), *Direct testing of speaking proficiency: Theory and application* (pp. 129-149). Princeton, NJ: Educational Testing Service.

American Psychological Association. (1985). *Standard for educational and psychological testing*. Washington, DC: Author.

Bachman, L. F. (forthcoming). *Fundamental considerations in language testing*. Reading, MA: Addison-Wesley.

Bachman, L. F., & Clark, J. L. D. (1987). The measurement of foreign/second language proficiency. *Annals of the American Association of Political and Social Sciences, 490*.

Bachman, L. F., & Palmer, A. S. (1981). The construct validation of the FSI oral interview. *Language Learning, 31*, 67-86.

Bachman, L. F., & Palmer, A. S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly, 16*, 449-465.

Bachman, L. F., & Savignon, S. J. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. *Modern Language Journal, 70*, 380-90.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*, 1-47.

Carroll, J. B. (1941). A factor analysis of verbal abilities. *Psychometrika, 6*, 279-307.

Carroll, J. B. (1963). The psychology of language testing. In A. Davies (Ed.), *Language testing symposium: A psycholinguistic approach* (pp. 46-69). London: Oxford University Press.

Carroll, J. B. (1983). Psychometric theory and language testing. In J. W. Oller (Ed.), *Issues in language testing research* (pp. 80-107). Rowley, MA: Newbury House.

Childs, J. (1987). Language proficiency levels and the typology of texts. In H. Byrnes & M. Canale (Eds.), *Defining and developing proficiency: Guidelines, implementations, and concepts* (pp. 97-106). Lincolnwood, IL: National Textbook Company.

Clark, J. L. D. (Ed.). (1978). *Direct testing of speaking proficiency: Theory and application*. Princeton, NJ: Educational Testing Service.

Clifford, R. T. (1978). Reliability and validity of language aspects contributing to oral proficiency of prospective teachers of German. In J. L. D. Clark (Ed.), *Direct testing of speaking proficiency: Theory and application* (pp. 191-209). Princeton, NJ: Educational Testing Service.

Cronbach, L. J. (1971). Validity. In R. L. Thorndike (Ed.), *Educational measurement* (pp. 443-597). Washington, DC: American Council on Education.

Cronbach, L. J. (1980). Validity on parole: How can we go straight? *New Directions for Testing and Measurement, 5*, 99-108.

Cronbach, L. J. (1984). *Essentials of psychological testing* (4th Edition). New York: Harper & Row.

Cronbach, L. J. (1985). *Construct validation after thirty years.* Paper given at the Symposium on Intelligence in honor of Lloyd Humphreys, Urbana, IL.

Dandonoli, P. (1987). ACTFL's current research in proficiency testing. In H. Byrnes & M.Canale (Eds.), *Defining and developing proficiency: Guidelines, implementations, and concepts* (pp. 75-96). Lincolnwood, IL: National Textbook.

Faerch, C., & Kasper, G. (1983). *Plans and strategies in foreign language communication.* In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 20-60). London: Longman.

Fitzpatrick, A. R. (1983). The meaning of content validity. *Applied Psychological Measurement, 7,* 3-13.

Guion, R. M. (1977). Content validity: The source of my discontent. *Applied Psychological Measurement, 1,* 1-10.

Hambleton, R. K. (1984). Validating the test scores. In R. A. Berk (Ed.), *A guide to criterion-referenced test construction* (pp. 199-230). Baltimore, MD: John Hopkins University Press.

Higgs, T. V., & Clifford, R. T. (1982). The push toward communication. In T. V. Higgs (Ed.), *Curriculum, competence, and the foreign language teacher* (pp. 57-79). Skokie, IL.: National Textbook Company.

Hymes, D. H. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269-293). Harmondsworth, England: Penguin.

Hymes, D. H. (1982). *Toward linguistic competence* [Mimeograph]. Philadelphia: Graduate School of Education, University of Pennsylvania.

Johnson, K. (1982). Some communicative processes. In K. Johnson (Ed.), *Communicative syllabus design and methodology.* Oxford: Pergamon.

Kane, M. T. (1982). A sampling model for validity. *Applied Psychological Measurement, 6,* 125-160.

Linn, R. L. (1980). Issues of validity for criterion-referenced measures. *Applied Psychological Measurement, 4,* 547-561.

Liskin-Gasparro, J. (1984). The ACTFL proficiency guidelines: A historical perspective. In T. V. Higgs (Ed.), *Teaching for proficiency: The organizing principle* (pp. 11-42). Lincolnwood, IL: National Textbook Company.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3,* 635-694.

Lowe, P., & Liskin-Gasparro, J. (undated). *Testing speaking proficiency: The oral interview.* ERIC Clearinghouse on Languages and Linguistics, Q & A. Washington, DC: Center for Applied Linguistics.

Lowe, P. (1978). Third rating of FSI interviews. In J. L. D. Clark (Ed.), *Direct testing of speaking proficiency: Theory and application* (pp. 159-169). Princeton, NJ: Educational Testing Service.

Madaus, G. F. (Ed.). (1983). *The Courts, validity, and minimum competency testing.* Boston, MA: Kluwer-Nijhoff.

Magnan, S. (1986). Assessing speaking proficiency in the undergraduate curriculum: Data from French. *Foreign Language Annals, 19,* 429-438.

Messick, S. A. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist, 30,* 955-966.

Messick, S. A. (1981a). Constructs and their vicissitudes in educational and psychological measurement. *Psychological Bulletin, 89,* 575-588.

Messick, S. A. (1981b). *Evidence and ethics in the evaluation of tests.* Research Report no. 1981b-81-89. Princeton, NJ: Educational Testing Service.

Mullen, K. A. (1978). Determining the effect of uncontrolled sources of error in a direct test of oral proficiency and the capability of the procedure to detect improvement following classroom instruction. In J. L. D. Clark (Ed.), *Direct testing of speaking proficiency: Theory and application* (pp. 171-189). Princeton, NJ: Educational Testing Service.

Mullen, K. A. (1980). Rater reliability and oral proficiency evaluations. In J. W. Oller & K. Perkins (Eds.), *Research in language testing* (pp. 91-101). Rowley, MA: Newbury House.

Oller, J. W. (1976). Evidence of a general language proficiency factor: An expectancy grammar. *Die Neuren Sprachen, 76*, 165-174.

Oller, J. W. (1979). *Language tests at school: A pragmatic approach.* London: Longman.

Oller, J. W. (1983). A consensus for the eighties? In J. W. Oller (Ed.), *Issues in language testing research* (pp. 351-356). Rowley, MA: Newbury House.

Popham, W. J. (1978). *Criterion-referenced measurement*. Englewood Cliffs, NJ: Prentice-Hall.

Savignon, S. J. (1972). *Communicative competence: An experiment in foreign-language teaching*. Philadelphia: Center for Curriculum Development.

Savignon, S. J. (1983). *Communicative competence: Theory and classroom practice.* Reading, MA: Addison-Wesley.

Stevenson, D. K. (1981). Beyond faith and face validity: The multitrait-multimethod matrix and the convergent and discriminant validity of oral proficiency tests. In A. S. Palmer, P. J. M. Groot, & G. A. Trosper (Eds.), *The construct validity of tests of communicative competence* (pp. 37-61). Washington, DC: TESOL.

Upshur, J. A., & Homburg, T. J. (1983). Some relations among language tests at successive ability levels. In J. W. Oller (Ed.), *Issues in language testing research* (pp. 188-201). Rowley, MA: Newbury House.

Upshur, J. A. (1979). Functional proficiency theory and a research role for language tests. In E. J. Briere & F. B. Hinofotis (Eds.), *Concepts in language testing: Some recent studies* (pp. 75-100). Washington, DC: TESOL.

Vollmer, H. J., & Sang, F. (1983). Competing hypotheses about second language ability: A plea for caution. In J. W. Oller (Ed.), *Issues in language testing research* (pp. 29-75). Rowley, MA: Newbury House.

Widdowson, H. G. (1978). *Teaching language as communication*. Oxford: Oxford University Press.

LEGAL CAVEATS ON COMMUNICATIVE PROFICIENCY TESTING FOR GRADUATION REQUIREMENTS OR TEACHER CERTIFICATION

Sally Sieloff Magnan
University of Wisconsin-Madison

1. Introduction

My remarks will focus on two concerns to which you call attention in your provocative paper: potential uses, and abuses, of proficiency testing, be it with the ACTFL/ETS/ILR interview (OPI) or another procedure, and the suggestion that we consider a multi-component system of scoring instead of a single, global score. Specifically, I will discuss two purposes for required testing of communicative proficiency: to demonstrate satisfactory completion of a language requirement (generally with regard to university entrance or exit) and to certify foreign language teachers (either as part of their undergraduate degree program or through a state mandate for licensing). For both these purposes, the OPI has frequently been selected or considered as the required testing instrument.

2. Language requirement

We may soon find considerable activity in creating proficiency requirements evaluated by a nationally-recognized test such as the OPI, since the Commission on Foreign Languages, Literatures, and Linguistics of the Modern Language Association has recently recommended that comparable institutions establish entrance and exit competency levels for all undergraduate foreign language students (Carduner, et al., 1986). Two major institutions already offer us models in this respect. The University of Minnesota is developing tests based on the OPI and the ACTFL Guidelines for use with a foreign language proficiency entrance standard for incoming freshmen in French, Spanish, and German (Arendt, Lange & Wakefield, 1986). The University of Pennsylvania has, since 1981, used the OPI as part of a battery of tests required to satisfy the language requirement in a number of foreign languages (Freed, 1984).

Data from the OPI indicate that students completing a college language requirement of one, or more likely two years, could be expected to rate within the intermediate level (Magnan, 1986). This is the level of minimum linguistic survival at which students are able to discuss aspects of their own lives in a simple, straight-forward fashion and can ask and answer simple questions both about their lives and in situations that they might encounter in the target community.

Kramsch (1986) and Schulz (1986) have cautioned that foreign language programs serving liberal arts requirements should not be limited to goals tested in the OPI, since this would likely be at the expense of, not only listening, reading, and writing, but also of international understanding and insights into how languages function as discourse systems. We should remember that tests are currently under development in listening and reading and that testing the speaking skill may bring speaking from a somewhat neglected status into its appropriate place in a balanced curriculum. We should also be careful that our enthusiasm for proficiency testing as a means of supporting our curriculum does not become a driving force that will misdirect or limit that curriculum. Bachman's suggestion that we investigate multi-component scoring could yield insights into the complexity of the abilities we wish to measure and thereby prove useful even if a global scoring procedure with its ease of use and interpretation proves more appealing to institutions and the general public.

When discussing testing for a language requirement, we must also consider issues of legality, as Bachman points out. A court case from 1979, mentioned by Bachman in his notes, deserves careful examination. In Debra P. versus Turlington (Florida State Board of Education), trial and appellate courts considered the validity of a test for functional literacy required for earning a high school diploma. While both courts ruled that the test had adequate content validity and thus was a valid measurement instrument, the appellate court ruled that the test was invalid for use in granting a high school diploma on the basis of instructional validity, that is, that the test was "fundamentally unfair in that it may have covered matters not taught in the schools of the state" (Madaus, 1983, p. 22). The Florida test was also found to discriminate on the basis of race in that black students graduating in 1979 had not had the benefit of fully equal schooling throughout their school careers.

Might not our future students, irate at being denied a college diploma because they fail a foreign language proficiency test, claim that their foreign language courses did not adequately prepare them for the required test? Might they not also question whether our proficiency test is biased toward a social class or ethnic group other than their own? As Liskin-Gasparro (1987) has pointed out, the Educational Testing Service experienced problems using the OPI with Peace Corps volunteers since the test seemed biased toward white collar professional language. We must include on our research agenda studies to determine if the OPI contains such an objectionable bias, particularly at the intermediate level of proficiency where our language requirements are likely to be set. Information provided by multi-component scoring, in addition to global scoring, could be useful in resolving potential challenges in this area.

3. Teacher certification

A second and even more immediate area of concern with regard to the use of tests of communicative proficiency is in teacher certification. Carroll's 1967 study and my 1986 findings indicate that foreign language majors are likely to score between intermediate high and advanced plus on the OPI scale. These levels span a wide range of ability, from that of the "good" requirement student already discussed, to that of the person who can handle most personal and professional situations, although with a certain degree of inconsistency. Perhaps because foreign language majors differ so greatly in ability, we clearly need a strong assessment instrument to ensure that we put teachers who are competent linguistically into our classrooms.

The time is now right to upgrade our teacher certification standards. According to a 1985 report from the Joint National Committee for Languages, the highest number of high school students are enrolled in foreign languages since 1968, and more than 60 universities have in the past few years reinstituted the foreign language requirement for graduation (Edwards, 1985). It is not surprising, therefore, that, as Clark and Clifford (1987) have indicated, an increasing number of states are reexamining teacher preparation standards and are considering requiring future teachers to demonstrate communicative language proficiency in the foreign language they are teaching.

There are three different bodies that might impose proficiency tests for teachers: the degree-granting institution, the hiring school district, or the professional teacher organization. If the test were to be imposed by the degree-granting institution, the test would need both content and instructional validity, as the Debra P. case shows. This requirement could potentially produce a limiting effect on either the test or the curriculum. If the hiring school district were to impose the test, the job relatedness of the test could be questioned. Another 1979 court case (Armstead vs. Starkville Municipal Separate School District) offers a precedent here: the graduate record exam was found unconstitutional for use in the hiring and retention of high school teachers, since it could not be proven that the tasks required in the exam were tasks essential to good teaching (Madaus, 1983, p. 246).

It would therefore seem most appropriate that, at least initially, proficiency tests for teacher certification be imposed, or suggested, under the auspices of the professional teacher organization. At least three bodies are now discussing language proficiency testing for foreign language teachers: the American Association of Teachers of French, the American Council on the Teaching of Foreign Languages, and the American Association of Colleges for Teacher Education. Of particular interest in response to Bachman's paper, is a resolution by this last body that calls for developers of standardized tests "to provide statistical information about their tests' reliability and validity, including an analysis of how ethnic and gender groups perform [...], extend truth-in-testing policies to such exams, [...] and abide by standards developed by the American Psychological Association for educational and psychological tests" (Olson, 1987). As Bachman points out, developers of the OPI would serve the profession well by obtaining this necessary information.

A good first step in research would be the study suggested by Clark and Clifford (1987) on the correlation between ratings on the OPI and actual performance of teachers, both inside the classroom and in non-classroom situations where teachers gain information about the target community. A combination of global and multi-component scoring could be used here to provide greater insights for validation of the current OPI procedure or for its eventual modification or replacement.

4. Conclusion

The problem we are addressing at this symposium is to come to terms with, as Byrnes (1987) has said, the tension between a testing model that is theoretically ideal and one that is practically feasible. Cronbach (1980) describes this tension well:

> Professionals who think of their work as scientific and objective are pained by references to consensus, persuasion, and plausible reasoning. The bad news is that we will not achieve unequivocal perfect evidence of validity. The good news is that no one expects or demands it, save as a rhetorical play. The world turns on judgmental weighing of uncertainties.

Our current problem is to decide how much uncertainty we can afford and where we should allow this uncertainty to be.

Bachman believes that the current OPI procedure may allow uncertainty to a dangerous degree and that two ways of increasing the validity of the interview would be to scrutinize carefully situations in which proficiency testing is to be used and to replace global scoring with multi-component scoring. While I am not convinced that global scoring cannot be valid for most of testing purposes and will not ultimately prove more useful, I do agree that a system of multi-component scoring should be investigated carefully, both to validate the global score and to provide further information in difficult decision cases.

We must be wary both of placing our profession in legal jeopardy and of putting ourselves into a position in which our desire to use a test will force us to make unwanted changes in our curriculum. As I see it now, our curriculum is being enhanced through influence from the OPI. The current interest in proficiency is leading us to greater understanding and more implementation of communicative teaching strategies. We now need to strive together to continue this positive direction and guard against future problems.

References

Arendt, J., Lange, D., & Wakefield, R. (1986). Strengthening the language requirement at the University of Minnesota: An initial report. *Foreign Language Annals, 19*, 149-156.

Byrnes, H. (1987). Features of pragmatic and sociolinguistic competence in the oral proficiency interview. This volume.

Carduner, J. (chair) et al. (1986). Recommendations of the Commission on Foreign Languages, Literatures, and Linguistics. *Profession 86*, Modern Language Association of America, 46-49.

Carroll, J. (1967). Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals, 1*, 131-151.

Clark, J., & Clifford, R. (1987). The FSI/ILR/ACTFL proficiency scales and testing techniques: Development, current status, and needed research. This volume.

Cronbach, L. (1980). Selection theory for a political world. *Public Personnel Management, 9*, 37-50.

Edwards, J. D. (1985). Catching up with the foreign language boom. *Wisconsin Association of Foreign Language Teachers Bulletin*, (Spring), 11-13. (Reprinted from the *New York Times*, December 19, 1985).

Freed, B. (1984). Proficiency in context: The Pennsylvania experience. In S. Savignon & M. Berns (Eds.), *Initiatives in communicative language teaching* (pp. 211-240). Reading, MA: Addison-Wesley.

Kramsch, C. (1987). From language proficiency to interactional competence. *Modern Language Journal, 70*, 366-372.

Liskin-Gasparro, J. (1987). The ACTFL proficiency guidelines: An update. This volume.

Madaus, G. (1983). Minimum competence testing for certification: The evolution and evaluation of test validity. In G. Madaus (Ed.), *The courts, validity, and minimum competence testing* (pp. 21-61). Boston: Kluwer-Nijhoff.

Magnan, S. (1986). Assessing speaking proficiency in the undergraduate curriculum: Data from French. *Foreign Language Annals, 5*, 429-438.

Olson, L. (1987). AACTE debates accreditation, releases new enrollment data. *Education Week, 6* (22), 1 & 18.

Schulz, R.A. (1986). From achievement to proficiency through classroom instruction: Some caveats. *Modern Language Journal, 70*, 373-390.

REACTIONS TO LYLE BACHMAN'S PAPER "PROBLEMS IN EXAMINING THE VALIDITY OF THE ACTFL ORAL PROFICIENCY INTERVIEW"

Elana Shohamy
Tel Aviv University

In reacting to Bachman's paper I would like to begin by stating that I agree with most of the points made in the paper, especially with regard to the need to validate the Oral Proficiency Interview (OPI), the research that needs to be undertaken and the far reaching and dangerous consequences of not doing so. I will elaborate on some of the points and add others which I deem to be important.

Construct validity may be the most elegant and strongest way to show the validity of a test. However, it is also the most difficult one to find evidence for. Therefore the measurement literature has suggested alternative ways for validation, such as CONTENT validity and CRITERION validity (concurrent and predictive), which are not only legitimate ways of validation but also ways which have some direct and practical merit for certain purposes. For example, if a test is used to predict certain oral language behavior in some future situation, then the extent to which the test actually predicts that type of oral behavior in that future, well defined context, will provide evidence for its predictive validity. Since the type of validity that needs to be pursued for a given test depends largely on its purpose and uses, it will be worthwhile for ACTFL to conduct an analysis of the purposes and intended uses of the OPI and to determine accordingly the type(s) of validity that need(s) to be researched. However, at the moment, there is no evidence of any type of validity for the OPI test, and I certainly agree with Bachman about the possible implications of using a test for which no validity has been demonstrated.

Although it is true that validity is the main issue when it comes to tests, reliability is an important issue as well. Bachman (in the abstract of the paper), states that the problem of the OPI is not related to reliability. I would like to claim, however, that reliability when it comes to the OPI is indeed a very important issue that cannot be overlooked, especially since it is a prerequisite for all types of validity. It is true that with extensive and continuous training of raters it is possible to obtain acceptable levels of inter-rater reliability for the OPI. (See for example, Hinofotis, 1976; Shohamy, 1983b); however, there are serious problems with the other types of reliability, specifically test re-test, inter-interviewer, and parallel forms reliability. As research has shown (Shohamy, 1983a) these types of reliability do pose a threat to the accuracy of the test taker's scores on an oral interview. For example, a change of the interviewer on an oral interview may affect the score s/he obtains on an oral interview. This seems logical since, as stated by Bachman and Savignon (1986), the tester is an integral part of the oral interview test whose personality, attitudes, rapport with the test taker, etc., may affect the quality of the language produced in the oral interview. The different types of reliability relevant to the OPI need to be examined and steps should be taken accordingly, to assure that the test does have high reliability. (For example, more strict standardization may be required in the event that inter-interviewer reliability is found to be low). Research on the reliability of the OPI must be conducted in order to determine the accuracy of the scores which test takers obtain on that test; that type of research needs to precede the research on the validity of the test.

Another issue which relates to the oral trait and the method used to tap the trait is the extent to which an oral interview is a representative sample of the oral trait. Specifically, viewing oral language as constituting a multiple of different speech styles and functions (e.g. discussing, arguing, apologizing, interviewing, conversing, being interviewed, reporting, etc.) means that being interviewed, the speech style and function tapped in an oral interview, represents only a single type of oral interaction. No doubt that it is an important speech style, but clearly, there are also other oral interactions which are equally important in real life situations.

In a number of research studies conducted in this area in recent years, it became clear that the performance of a test taker in one type of oral interaction, an interview, cannot accurately predict the test taker's performance on another type of oral interaction, a reporting task. (Shohamy, 1983a). In a more recent study (Shohamy, Reves & Bejerano, 1986), it was found that the correlations among four different oral interactions, an oral interview, a reporting task, a role play, and a discussion, were only moderately correlated with one another. Clearly, an oral interview, by itself, cannot be a valid predictor and measure of a test taker's oral proficiency. It seems, therefore, that in order to tap oral proficiency in a more valid way, there is a need to administer a number of oral tests, based on a number of oral interactions and thus to obtain a more representative sample of oral proficiency. (Preferably each conducted by a different tester).

The above research led the Ministry of Education in Israel to introduce a new national oral proficiency test in English as a Foreign Language, instead of the oral interview which had been in use for over 20 years. In this new test each student (at the end of 12th grade) is being tested by a number of different oral interactions: an interview, a monologue, a role play and a group discussion, each by a different tester. (See Shohamy et al., 1986, for details on the research and the implementation of the test). It may be worthwhile for ACTFL to consider such a model and to try to test oral proficiency via different types of oral interactions and thus to obtain a more representative sample of oral proficiency. Interestingly, a by-product of such a test is its strong backwash effect on instruction, i.e. in language classes students are taught different oral interactions.

The last point I will make relates to the need to go 'beyond validity'. Although validity is a necessary criterion for judging the quality of tests, it is not a sufficient one, especially when it comes to the OPI. Nevo and Shohamy (1986), call for a focus on additional aspects beyond reliability and validity, and to view tests in the larger educational, social and political context in which they operate. Specifically, we propose a focus on UTILITY, used to ensure that a test will serve practical information needs of a given audience; ACCURACY, intended to assure that a test will reveal or convey technically adequate information on the educational achivement of those being tested; FEASIBILITY, intended to ensure that a test will be realistic, prudent and frugal; and FAIRNESS, intended to ensure that the test is conducted legally, ethically, and with regards to the welfare of the tested individuals as well as those affected by the test results.

Specifically in the case of the OPI which was introduced for reasons that go beyond assigning oral test scores to students, i.e. to create a positive backwash effect on instruction and thus to upgrade the achievement and proficiency in foreign language in the USA. There is a need to evaluate that test by these other purposes as well.

In summary, 1) there is need to examine different types of validity as well, such as content and criterion, in line with the purpose of the test. 2) there is a need to examine the different types of reliability for the OPI test, in order to assure the accuracy of the test taker's scores. There are still some serious problems of reliability with the oral interview test. 3) there is a need to consider other speech styles and functions of oral interaction in addition to "an interview", in order to obtain a more representative and valid sample of oral proficiency. Finally, 4) there is a need to go beyond reliability and validity towards utility, feasibility and fairness of the test, and thus to judge the OPI test in the educational, social and political context in which it had been introduced.

For all these needs massive research and development is urgently called for, the results of which will extend our understanding of the OPI test and its uses. Without such information the test is too dangerous a weapon.

References

Bachman, L. F., & Savignon, S. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. *Modern Language Journal, 70*, 380-390.

Hinofotis, F. (1976). *An investigation of the concurrent validity of cloze testing as a measure of overall proficiency in English as a second language.* Unpublished doctoral dissertation. Southern Illinois University.

Nevo, D., & Shohamy, E. (1986). Evaluation standards for the assessment of alternative testing methods: An application. *Studies in Educational Evaluation, 12*, 149-158.

Shohamy, E. (1983a). The stability of oral proficiency assessment on the oral interview testing procedure. *Language Learning, 33*, 527-539.

Shohamy, E. (1983b). Rater reliability of the oral interview speaking test. *Foreign Language Annals, 16*, 219-222.

Shohamy, E., Reves, T., & Bejerano, Y. (1986). Introducing a new comprehensive test of oral proficiency. *English Language Teaching Journal, 40*, 212-222.

Summary of the discussion session on PROBLEMS IN EXAMINING THE VALIDITY OF THE ACTFL/ETS PROFICIENCY INTERVIEW

Cheryl Engber
Indiana University

Following the presentation by Lyle Bachman and responses by Sally Magnan and Elana Shohamy, discussion centered on three general areas of concern: theoretical issues, pedagogical issues, and public/political issues. Perhaps because Bachman's paper highlighted the problem of testing the validity of the ACTFL/ETS oral interview along with the related theoretical question of reliability, much of the discussion focused on these issues as well.

The question was raised as to the content validity of the oral interview. Two issues that seemed to be relevant to this point were discussed. A participant suggested that the first issue could be called "procedural validity"; he questioned whether the procedures which trainers use actually lead to validity in its various senses. Of particular concern is the fact that the oral interview is a one-item type test. The suggestion was made that including other types of tasks such as monologues, the use of various question types to elicit performance, and role-plays would enhance the procedural validity of the test.

A second, related point was made with regard to the practice of defining the scales of the test in terms of language functions. This practice poses the great dilemma of content validity. Because a test must be viewed as a sample of performance, it is incumbent on the tester to demonstrate that the sample one gets is, in fact, a representative sample of performance. As the participant pointed out, however, no exhaustive list of language functions exists for any language. Thus, the dilemma arises that as long as the scale levels are defined in terms of actual language functions and actual language performance, there is absolutely no way of demonstrating content validity.

Finally, with regard to validity, it was suggested that the orientation should be in the direction of viewing validity in a macro-sense. In other words, to guard against falling into the trap of saying that a test is valid or invalid, the test should be seen as valid if it works for a particular purpose. Thus, every time there is another purpose, there is a different kind of validity as well as a different way of exploring and explicating that validity. Investigation into validity would then center on how one knows that a test works well for a particular purpose yet works differently for some other purpose.

Besides the problem of demonstrating some kind of validity for the oral interview, the related problem of reliability was also discussed. Of particular concern was the issue of inter-rater reliability. The question was asked about the nature of the social construct of the consensus of a group of testers who are socialized into being good testers. Is the construct that of native speakers or non-native speakers who know the form of the language and culture, or is it an American consensus? Furthermore, when the testers test, do they put on a native-speaker view/perception of the testee, or do they put on an American tester personna or a teacher's personna? The response given to these questions came from an academic standpoint. It was noted that testers have been trained to match their impressions at the time against impressions they have been trained to use as models during the training sessions. In effect, impressions are matched against the descriptions at different levels. These descriptions have evolved from listening to native speakers, from

DLI observations of people in various job categories, and from lots of models of students whose oral production reflects a range of language proficiency.

The response to this point was that, nevertheless, some concern remained that native speaker reactions to error vary widely. In addition, the situation may be more complicated by the fact that what a tester does at the lower end of the scale may, in fact, reflect an American bias. This is especially true since the system was originally developed to see how well Americans use foreign languages. As one moves up the scale, more native-like ways of looking at things are assumed, and one of the decisions made is whether the speaker will be accepted as a native speaker. While the latter may be true of the government test, it may not be as crucial when using the academic scale, according to one respondent. Because the superior levels are rarely needed in the academic world, testers don't have to deal with how close a speaker approaches the native speaker norm. It is only necessary to be able to say that a person can perform all of the functions associated with a particular level. (This use of the notion of functions, as was noted earlier, presents problems for demonstrating the validity of the test.)

Continuing the discussion of reliability, one participant noted that the question still remains with regard to how reliably testers agree (whatever the nature of the task may be at the different levels.) Although they did not elaborate, government spokespersons mentioned that the government has done several inter-rater reliability studies. On the academic side, not much has been done, but ACTFL has submitted a proposal to the Department of Education to do such a study.

The government has also done at least one test/re-test reliability study. It consisted of a semi-direct test with Americans learning Chinese. Of four forms of the test, students took two of the forms in addition to a live interview. Test/re-test reliability was found to be in the low nineties. The government also conducts intra-reliability studies every year. These studies are intended to check the reliability of an individual tester against himself/herself over time. It was suggested that every two years would be adequate for the academic test.

In addition to validity and reliability, the issue of global scoring generated a great deal of discussion -- some more theoretical, some more pedagogical. Bachman clearly pointed out the need for a multi-componential rating. This was questioned by at least one discussant who noted that in the academic world, one grade is always given in a multi-dimensional course. Furthermore, students are admitted to graduate school at least partially on the basis of GPA which is an accumulation of all these single grades. (While it was claimed that the SAT, GRE, and TOEFL tests are one-score tests, this was quickly disputed at least for the SAT and GRE.) On a theoretical level, it was noted that the multi-componential score is needed if for no other reason than to be able to ratify the global score.

Pedagogically, such a rating would allow for recoverability of components for the students. Thus, the interview and the resultant multi-componential rating would serve a diagnostic purpose. It could also serve as the basis for a better understanding of the obviously related but still differentially perceptible components of language processing that go into a student's ability at any rating level. To further emphasize this point, it was mentioned that even though a single score may have been used traditionally, merely substituting numbers (as the ACTFL/ETS test does) is not a leap forward for the language teaching profession and thus should not be justified on that basis. Even the fact that the one score is a compilation of several sub-elements does not provide for the recoverability of these components for the student.

Elana Shohamy, in her response to Bachman's paper, was particularly concerned about this need to be able to provide feedback to students. One participant, commenting on this area of concern, acknowledged that one global rating even with the accompanying verbal description is not as useful as being able to sit down and break out that rating into the component parts, a procedure which can and should be done pedagogically on a one-to-one basis. Another participant extended this response by pointing out that reality is on the side of informal testing. The rating itself may be almost irrelevant for diagnostic purposes. However, on the basis of the criteria that the rating scale does incorporate, an instructor can talk to a student and clarify what they do well and point out areas where they need to make progress.

Thus, for diagnostic pedagogical purposes, a multi-componential rating would be useful. Discussion returned, nevertheless, to the need for one number for standardizing purposes. One number provides a clear, understandable, standardized measure which is acceptable to the general public when it comes to decision making. In response, it was emphasized that the fact that decision makers only look at a single score doesn't absolve test-developers from the responsibility of providing a full complement of relevant information. If the user ignores the information and only looks at the single score, it is beyond the control of the test-developer.

Finally, then, it was emphasized that the issue was not to do away with the single global score but to add a multi-componential rating for several reasons: (1) for validation purposes, (2) for diagnostic purposes, and (3) for providing a back-up for the tester and other decision makers should the one number be questioned. Further, what is needed is research which will look into and match the single score against a componential rating (once a decision has been made as to what the components will be) before deciding what use to make of the multi-componential rating - diagnostic, validation of the global score, back-up for the global score, or to replace the global score (although the participant disagreed with the latter.)

The role of a multi-componential rating as a back-up for the global score impinges somewhat on a final portion of the discussion - that of the political ramifications of the use of the oral proficiency interview. This is of particular concern to the academic world as the test is being used more and more widely for teacher certification and similar purposes. It was strongly suggested that it is necessary to differentiate between classroom testing and the one-shot test such as the oral proficiency interview. While classroom testing can be dangerous, the sample is much larger, and there are different kinds of tests. Thus, according to the participant, classroom testing is not nearly as dangerous as the one-shot test, especially, as has been noted, since there is no real knowledge about its validity, and the test is being used to dictate the lives of people.

One case of litigation was mentioned. The oral proficiency interview has been (and still is) used in New Jersey for bilingual certification. In 1978 or 79, two bilingual teachers sued the State Education Department because they claimed that (1) bilingual teachers shouldn't be singled out for special treatment, i.e., they claimed to be the only classroom teachers who were required to show competency in their field, and (2) the test was capricious, invalid, and unreliable. The federal courts found for the state by deciding that while there were procedural flaws, the test was not an invalid or unreliable one. The state was aided in its defense by the fact that the decisions were never irrevocable (people had the opportunity for a re-test), the test was anonymous, recorded, and multiply rated; in addition, the judge listened to a sample and questioned how the teachers could even function in English.

As noted by Sally Magnan in her presentation, this ending is not always assured. She cited a case in which a proficiency exam was used to withhold a diploma from a black student. The courts ruled that the test may have had content validity but had no instructional validity. It may, in fact, have covered things not taught in the schools and/or was premised on racial discrimination because blacks had not had the same educational opportunities as whites. Guarding against litigation, then, must be considered as the oral proficiency interview comes into more widespread use; validity cannot be ignored as that will be the central issue in any court case.

Thus, a degree of consensus was reached during the discussion with regard to the need for research into various kinds of validity, reliability, and a possible multi-componential rating. A final question was then addressed to Lyle Bachman. The questioner agreed that it would be nice to be able to demonstrate to the outside world the basis for the oral proficiency interview, but since Bachman does not seem to want to see function, content, context, and accuracy used as the basis for the criterial statements, she wondered how test-developers could come up with a hierarchical statement that would get at the different levels of oral proficiency. He replied that all of the things now in the scale are empirical questions, subject to empirical investigation. One possible way to investigate from a qualitative point of view would be to look at hundreds of oral interview tapes and analyze the language functions. This would serve mainly to find the in-between levels. The empirical test would be to see how many levels raters could reliably identify. He stressed that a theoretical framework is necessary in order to begin defining components. In other words, the definitions have to be treated as hypotheses.

When questioned further whether the rating scale as it exists could serve as a hypothesis, he responded that it could but that there are problems. The most important is the confounding of content, function, specific tasks, interactions between the tester and test-taker, and others. In particular, content, functions, and tasks are confounded because they are not well defined. One could see to what extent these could be pulled from the existing scale and then design studies to look at them. If we could define these, they could be looked at empirically.

From this discussion, it is obvious that many research projects are possible. Devising a multi-componential rating and comparing it with the global score seems advisable. Particularly desirable would be studies which attempted to separate the various components included in the oral proficiency interview for empirical testing leading to definitions or criterial statements with an experimental rather than an experiential basis. Reliability studies of various kinds - inter-rater, test/re-test, intra-rater - are needed in the academic world. And perhaps most crucial are research efforts that would lead to a test which could be validated before it is used to affect the lives of not only the test-takers but the test-developers as well.

THE EXTENSION OF LANGUAGE PROFICIENCY GUIDELINES AND ORAL PROFICIENCY TESTING TO THE LESS COMMONLY TAUGHT LANGUAGES

David Hiple
American Council on the Teaching of Foreign Languages

My task is to address the application of proficiency guidelines and the oral proficiency interview to less commonly taught languages. For discussion purposes, I would like to draw a distinction among commonly taught languages, less commonly taught languages, and what I will call much less commonly taught languages. For the sake of comparison, let us review some enrollment figures for various foreign languages studied in the United States. The 1983 MLA Survey of Foreign Language Registrations in US Institutions of Higher Education indicates that 784,515 college students enrolled for courses in the three most commonly taught languages: 386,238 in Spanish, 270,123 in French, and 128,154 in German. Russian was the fifth most commonly taught language at the postsecondary level, but it was significantly less commonly taught than Spanish, since only 30,386 college students enrolled in Russian courses in 1983, less than ten percent of the number of students enrolled for Spanish.

Enrollment figures in 1983 for the first cluster of less commonly taught foreign languages at the college level (following Spanish, French, and German) were as follows: Italian 38,672; Russian 30,386; Hebrew 18,199; Japanese 16,127; Chinese 13,178; Portuguese 4,447; and Arabic 3,436. The number of students enrolled in Japanese courses in 1983 represented approximately four percent of the number of students enrolled in Spanish courses, and the number of students enrolled in Arabic courses represented approximately one percent of the number enrolled in Spanish courses.

After this cluster of less commonly taught languages, enrollments reflect that most other foreign languages are much less commonly taught; for example, in 1983, 507 college students enrolled for Swahili, 219 college students enrolled for Hindi, and 85 college students enrolled for Indonesian. The number of students enrolled in Indonesian courses represented approximately two percent of the number of students enrolled in Arabic courses. Yet, even Indonesian scholars, with their 85 students, could take comfort when they compared their students to the 14 students enrolled in Uzbek or the four students enrolled in Ibo.

At the secondary level, differences between enrollments in commonly taught languages and less commonly taught languages are even more dramatic. The 1982 ACTFL Survey of Foreign Language Enrollments in Public Secondary Schools indicated that Spanish, French, German, Latin, and Italian accounted for approximately 99% of the 2,740,198 foreign language enrollments. Of the remaining one percent, 5,497 students were enrolled in Russian, 1,980 were enrolled in Chinese, and 51 were enrolled in Arabic.

Thus, within what is generally referred to as the "less commonly taught languages," there is a tremendous range in student enrollments, reflecting the fact that some languages are "much less commonly taught." ACTFL's language proficiency activities have progressed from projects in "commonly taught languages," initiated in 1981, to projects in "less commonly taught languages," initiated in 1985. The initial projects involved proficiency guidelines for French, German and Spanish as well as training individuals to administer and evaluate oral proficiency tests in Spanish, French, German, and Italian. The second stage of activities involved proficiency guidelines for Chinese, Japanese, and Russian, and oral proficiency tester training in Arabic, Chinese, ESL/EFL, Japanese, Portuguese, and Russian. The third stage of activities involved a dissemination project, undertaken jointly with the Center for Applied Linguistics for Arabic, Hindi, Indonesian, and Swahili, as well as preliminary oral proficiency tester training activities in Hindi, Indonesian, Swahili and a small sample of other African languages such as Hausa.

1. Proficiency guidelines for the less commonly taught languages

The 1986 ACTFL Proficiency Guidelines represent a hierarchy of global characterizations of integrated performance in speaking, listening, reading, and writing. Each description is a representative, not an exhaustive, sample of a particular range of ability, and each level subsumes all previous levels, moving from simple to complex in an "all before and more" fashion. The introductory statement reminds readers that the construction and utilization of language proficiency guidelines is a dynamic, interactive process and that the academic sector, like the government sector, will continue to refine and update the criteria periodically to reflect the needs of the users and the advances of the profession. In terms of the extension of the 1982 Provisional Proficiency Guidelines to the less commonly taught languages, there was no doubt of the need to refine and update the criteria.

In 1983, ACTFL received support from the Department of Education to initiate the second stage of the guidelines project to create language-specific proficiency statements in Chinese, Japanese, and Russian. As the working committees began the task, the Indo-European bias of the existing generic guidelines became increasingly problematic. Each level description of the guidelines reflects the functional trisection and contains function statements, context statements, and accuracy statements. The Western bias was most evident in statements that addressed the accuracy component of the trisection. After completing the initial draft of the Chinese, Japanese, and Russian statements, it became evident that creating meaningful guidelines for those languages would not be possible without a revision of the generic guidelines. Thus, ACTFL petitioned the Department of Education and was granted an amendment to the project to revise the generics guidelines so as to create, in turn, viable guidelines for Chinese, Japanese, and Russian.

Let us look at the evolution of the proficiency guidelines as related to less commonly taught languages, tracing the development of the Speaking description of one level, Intermediate-Mid (I-M), from the Interagency Language Roundtable (ILR) Level Definitions through the Provisional Proficiency Guidelines to the 1986 Proficiency Guidelines and the respective language-specific descriptions in Chinese and Russian.

The first thing one notices about the ILR S-1 Definition is the orientation toward travel and the concomitant adult content references to asking for lodging, for example. For the academic Intermediate (ILR S-1) speaker, such content areas may be inappropriate; while the Intermediate speaker may be expected to satisfy survival needs, the academic descriptions add that the Intermediate speaker may be able to satisfy the level requirements in the content area of social situations or social demands; e.g., inviting a friend to a party, asking him/her to bring other friends and/or refreshments, and giving directions to the party.

Another obvious difference between the ILR S-1 definition and the academic Intermediate descriptions is that the equivalent to the ILR S-1 is represented by two sublevels, Intermediate-Low and Intermediate-Mid, on the academic scale. Liskin-Gasparro (4) recounts the ETS Common Yardstick Project and need for a scale that discriminates more finely at the low end, since academic students tend to cluster there. This need is particularly real in less commonly taught languages where students can expect to invest much more time in learning the target language before arriving at the Intermediate level. For example, the School of Language Studies of the Foreign Service Institute estimates that students may require twice as much time to attain S-1 proficiency in Arabic, Chinese, Japanese, and Russian than to attain the same level of proficiency in French or Spanish. Thus, the need to distinguish among subranges of Intermediate level proficiency seems particularly compelling for a number of less commonly taught languages.

ILR Skill Level Descriptions.

Elementary Proficiency (S-1). Able to satisfy routine travel needs and minimum courtesy requirements. Can ask and answer questions on very familiar topics within the scope of very limited language experience, can understand simple questions and statements, allowing for slowed speech, repetition or paraphrase; speaking vocabulary inadequate to express anything but the most elementary needs; errors in pronunciation and grammar are frequent, but can be understood by a native speaker used to dealing with foreigners attempting to speak the language; while topics which are "very familiar" and elementary needs vary considerably from individual to individual, any person at the S-1 level should be able to order a simple meal, ask for shelter or lodging, ask and give simple directions, make purchases, and tell time.

***ACTFL Provisional Proficiency Guidelines*:**

Intermediate-Low. Able to satisfy basic survival needs and minimum courtesy requirements. In areas of immediate need or on very familiar topics, can ask and answer simple questions, initiate and respond to simple statements, and maintain very simple face-to-face conversations. When asked to do so, is able to formulate some questions with limited constructions and much inaccuracy. Almost every utterance contains fractured syntax and other grammatical errors. Vocabulary inadequate to express anything but the most elementary needs. Strong interference from native language occurs in articulation, stress, and intonation. Misunderstandings frequently arise from limited vocabulary and grammar and intonation. Misunderstandings frequently arise from limited vocabulary and grammar and erroneous phonology but, with repetition, can generally be understood by native speakers in regular contact with foreigners attempting to speak their language. Little precision in information conveyed owing to tentative state of grammatical development and little or no use of modifiers.

Intermediate-Mid. Able to satisfy some survival needs and some limited social demands. Is able to formulate some questions when asked to do so. Vocabulary permits discussion of topics beyond basic survival needs such as personal history and leisure-time activities. Some evidence of grammatical accuracy in basic constructions, for example, subject-verb agreement, noun-adjective agreement, some notion of inflection.

The revision of the provisional guidelines in 1986 permitted many refinements. In the case of Intermediate-Low and Intermediate-Mid, the functional ability of speakers at both levels is essentially the same, thus the distinction is made in terms of quantity, the number of content areas, and quality, the degree of accuracy, sustained in speech. In the case of quantity, there was a recognized need to flesh out the Intermediate-Mid description and provide more examples of generic content areas representative of the level. In terms of quality, the problem was more serious, since too many of the accuracy requirements were reflective of Indo-European languages and irrelevant to Chinese and Japanese, for example. An attempt was made, therefore, to remove many of the quality statements and reserve them for their proper place in the language-specific guidelines.

A related refinement was the reformatting of the guidelines to present the generic and language-specific statements together so that a language-specific level description could support a generic level description and the two could be read simultaneously. This reformatting was particularly useful in light of the attempt to maintain the neutrality of the generic descriptions and to focus the accuracy statements in the language-specific descriptions.

In the reformatting, a summary of the hallmark characteristics of each major level, Novice, Intermediate, Advanced, and Superior, was also introduced in each skill area to reduce the density of the guidelines text. This new format was utilized in the creation of the Chinese and Russian guidelines, and, if it proves to be workable, all of the language-specific guidelines will at some point be reformatted similarly.

ACTFL Proficiency Guidelines, 1986:

Intermediate. The Intermediate level is characterized by the speaker's ability to:

-create with the language by combining and recombining learned elements, though primarily in a reactive mode;

-initiate, minimally sustain, and close in a simple way basic communicative tasks; and

-ask and answer questions.

Intermediate-Mid. Able to handle a variety of uncomplicated, basic and communicative tasks and social situations. Can talk simply about self and family members. Can ask and answer questions and participate in simple conversations on topics beyond the most immediate needs; e.g., personal history and leisure-time activities. Utterance length increases slightly, but speech may continue to be characterized by frequent long pauses, since the smooth incorporation of even basic conversational strategies may continue to be strongly influenced by first language and fluency may be strained. Although misunderstandings still arise, the Intermediate-Mid speaker can generally be understood by sympathetic interlocutors.

One other major decision in the 1986 revision of the proficiency guidelines and the accompanying guidelines for less commonly taught languages was the decision to suspend work on culture statements until such time that a more appropriate model for measuring cultural proficiency is developed. Galloway (1986, p. 70) evaluates the ACTFL experiment of the provisional culture guidelines by pointing out that they were "somewhat artificially forced into an inappropriate mold, and they became confined to a series of learned behaviors, wedded to the content/context ranges of the language progression...While the language guidelines reflect an assumption of a progression of increasingly complex linguistic and cognitive skills, the culture guidelines reflect no such hierarchical skills arrangement." Allen (1985, p. 149) summarizes, noting that "a new conception of cultural proficiency must replace that now in existence."

2. Oral proficiency testing in the less commonly taught languages

ACTFL currently conducts academic oral proficiency testing workshops in ten

Chinese. Can ask and answer simple questions involving areas of immediate need (*Yíge rén de fángjiān, yìtiān duōshǎo qián?*), leisure time activities (*Nǐ xǐhuan kàn wàiguó diànyǐng ma?*), and simple transactions, i.e., at the post office (*Hángkōngxìn duōshǎo qián?*). Quantity of speech is increased and quality of speech is improved. Greater accuracy in word order, basic constructions, and use of high-frequency verbs and auxillaries (*shì, yǒu, yào, xǐhuan*). Can use simple time words (*míngtiān, xiànzài*) and other time markers to indicate various time relations. Has basic knowledge of differences among such sets of terms as: *huì, kěyǐ, néng; jiào, qǐng, wèn; zhīdao, rènshi*, but still makes errors.

Russian. Speaker can sustain very short conversations by giving information about self (Я родился/родилась в ...; Мне ... лет; Я учусь в университете на ... курсе, на ... факультете); asking simple questions (Откуда вы? Где вы сейчас живете? Где вы учитесь?). Can talk simply about leisure time activities (Я люблю играть в ... ; Я собираю ...; Я увлекаюсь ...). Speaker can handle an increasing number of simple situations, such as grocery shopping (Дайте, пожалуйста, двести грамм ...); getting a hotel room (У вас есть номер на двоих? В номере есть ванная?); ask for and give directions (Вы не скажете, как лучше дойти до станции метро? Вам лучше ехать на метро); make simple arrangements (Приходите в шесть часов; Я вернусь через полчаса); ask for simple information about events (Когда начинается вечерний сеанс? У вас есть билеты на «Лебединое озеро»? Дайте, пожалуйста, два билета на восемнадцать часов); make simple travel arrangements (У вас есть билеты на утренний рейс до Киева? Дайте два билета на двадцатое декабря). Speaker can express some feelings and states (Мне холодно; Я люблю теплую погоду; Я очень люблю современную музыку). There is improving accuracy in adjective-noun and subject-predicate agreement, and in control of some cases, e.g., nominative, genitive, accusative and prepositional. There is partial control of some common imperatives (Скажите, дайте, идите, принесите, покажите). Able to use such verbs as хотеть, мочь, любить, though without complete control of conjugated forms.

languages: Arabic, Chinese, ESL/EFL, French, German, Italian, Japanese, Portuguese, Russian, and Spanish. As I have already stated, ACTFL's proficiency activities were initiated in the commonly taught languages, were gradually expanded to the less commonly taught languages, and are now being expanded further into what I am calling the much less commonly taught languages. In launching an academic oral proficiency testing program, ACTFL benefitted greatly from the input and experience of the ILR and of the CIA Language School and the Defense Language Institute (DLI), in particular. These two agencies lent personnel to ACTFL to conduct the initial training sessions in various languages, and as good academic testers emerged, they were given advanced training so as to become workshop trainers themselves. Thus, the dependency on government personnel has reduced gradually.

ACTFL has relied on CIA and DLI, rather than FSI, for personnel for two basic reasons. First, the interview as it is currently administered at FSI is somewhat more job specific and professionally oriented than the interview that is administered at CIA and DLI. The FSI test consists of three parts: warm-up conversation, examinee interviewing examiner on a topic of the examinee's choice, and examinee briefing examiner on selected topics and responding to comments/questions/challenges. The CIA/DLI interview, on the other hand, is somewhat less controlled in content and more conversational in structure. Second, CIA/DLI testers tend to be more global in their evaluation of the speech sample, whereas FSI testers are somewhat more discrete in their rating. Wilds (1975, p. 33) summarizes the differences, noting that "CIA testers tend less to analyze the examinee's performance in detail; functional effectiveness is the overriding concern." The end result is that the CIA/DLI interview is more easily applied to academic testing contexts.

In terms of extending the oral proficiency interview to additional less commonly taught languages, a major problem presents itself when no CIA/DLI tester is available to train academic testers in a particular language. ACTFL first addressed this problem in the case of Hindi. The solution, a time consuming one, is to train testers in a language other than the target language and subsequently to assist the most interested testers in transferring the concepts, procedures, and rating criteria to the target language. In addition to Hindi, transfer projects are currently in various stages of development in Hebrew, Indonesian, and Swahili.

In the case of Hindi, testers started the transfer by experimenting with the interview in the target language and gathering speech samples. The testers identified the best Hindi interviews, translated the questions asked in those interviews into English and met with English testers to get feedback on elicitation technique and assistance in rating samples. In rating, it was necessary to make some accommodation for Hindi-English code switching. Interlanguage is generally an indication of a relatively low level of proficiency, but in the case of Hindi, appropriate Hindi-English code switching also seems to be representative of educated native Hindi. Kahane (1986, p. 495) addresses this phenomenon in his discussion of prestige languages, indicating that "one of the primary motivations for acquiring the prestige language is its identification with education, which transfers it to the values of a class symbol."

The pioneer Hindi tester, Vijay Gambhir at the University of Pennsylvania, who is also a certified tester in English, has submitted a proposal to the Department of Education to conduct a computer analysis of the presence/absence of identified accuracy criteria at various proficiency levels, to interview educated native speakers of Hindi so to gather data on the extent to which Hindi-English code switching is characteristic of educated spoken Hindi, and to form a committee to write proficiency guidelines for Hindi.

In Arabic, another interesting problem presented itself as Arabists grappled with the relationship of diglossia to oral proficiency testing and proficiency guidelines. Roger Allen and the University of Pennsylvania have received a grant from the Department of Education to form a committee to establish proficiency guidelines in Arabic. Allen (1984, p. 121) points out that there are extreme differences in colloquial dialects within the Arabic-speaking world. He also points out that while Modern Standard Arabic is the only language used throughout the Arabic-speaking world, it is not the means of oral communication between any two speakers of one dialect area. The working solution for oral proficiency testing in Arabic is for the tester to conduct the interview in Modern Standard Arabic and to accept responses in Modern Standard Arabic or in any colloquial dialect. Only at the Superior level, the professional level, would the examinee be expected to demonstrate proficiency in Modern Standard Arabic. For the most part, students of Arabic in the United States learn Modern Standard, so to date little confusion has resulted from this solution.

Initiatives and issues in the much less commonly taught languages – In 1985 CAL and ACTFL undertook a collaborative Proficiency Dissemination Project funded by the Department of Education to extend the proficiency concept to targeted less commonly taught languages, including Arabic, Hindi, Indonesian, and Swahili.

Consultants representing the targeted languages met to summarize the state of the art vis-a-vis proficiency in their particular language and to suggest project activities to address expressed needs in their individual languages. Advisory consultants representing Chinese, Japanese, and Russian, where significant proficiency-oriented activities were already occurring, also participated in the project to share with the "new" language consultants the experiences of transferring proficiency-oriented activities in their particular languages.

A variety of activities has grown out of this project. Over the last two years, the target language consultants, the advisory consultants, and the project directors have given familiarization workshops on oral proficiency testing, including language-specific demonstrations in Arabic, Hindi, and Indonesian, at the conferences of such organizations as the Association for Asian Studies and the Middle Eastern Studies Association. In the next few months additional workshops, addressing proficiency based-instruction, will also be presented. In addition, print materials featuring contributions from the project consultants on the application of proficiency concepts to their particular languages will be distributed among the target language groups.

The priorities of the Department of Education and the initiatives of numerous groups and projects, such as the CAL/ACTFL project, occurring concurrently on many fronts have raised the level of awareness about proficiency-oriented teaching and testing in the less commonly taught languages and the much less commonly taught languages. At the same time, a number of related issues have come to the fore, and certain groups are now beginning to discuss ways to deal with these issues. The ten Title VI African Studies Centers, for example, have scheduled a meeting in May (1987) to examine ways to apply proficiency-oriented teaching and testing to much less commonly taught African languages and to prepare a position paper on proficiency applications to their languages, and the Title VI South Asia Centers met last month to discuss proficiency policy issues for their centers. The following are a few of the issues I have heard raised by educators in the less commonly taught languages:

I. How can proficiency testing and the development of proficiency guidelines be economically and practically applied to less commonly taught languages, such as Hausa?

II. Should proficiency guidelines be developed for every language?

III. Are there unique features of certain less commonly taught languages that resist measurement by a common metric?

IV. What modification of standards and procedures might be necessary in preparing oral proficiency testers in less commonly taught languages?

V. How does the average student of Spanish, for example, differ from the average student of Hindi, for example, and what special considerations might be taken into account when applying proficiency-oriented teaching and testing to a different population?

It is not my intention to offer solutions to these issues but rather to call attention to them and to propose them as a departure point for discussion. While many issues remain to be resolved, I believe that the extension of proficiency guidelines and oral proficiency testing to the less commonly taught languages has had a general positive effect not only on those working in the less commonly taught languages but also on those working in the commonly taught languages. The search for a common metric has encouraged new collaborations, and we have learned that we may have more in common than we thought.

References

American Council on the Teaching of Foreign Languages. (1982). *ACTFL provisional proficiency guidelines.* Hastings-on-Hudson, NY: Author.

American Council on the Teaching of Foreign Languages. (1986). *ACTFL Chinese proficiency guidelines.* Hastings-on-Hudson, NY: Author.

American Council on the Teaching of Foreign Languages. (1986). *ACTFL proficiency guidelines.* Hastings-on-Hudson, NY: Author.

American Council on the Teaching of Foreign Languages. (1986). *ACTFL Russian proficiency guidelines.* Hastings-on-Hudson, NY: Author.

Allen, W. (1985). Toward cultural proficiency. In A. C. Omaggio (Ed.), *Proficiency, curriculum, articulation: The ties that bind* (pp. 137-165). Middlebury, VT: Northeast Conference on the Teaching of Foreign Languages.

Allen, R. M. A. (1984). Proficiency guidelines for Arabic. *Al-'Arabiyya, 17*, 119-127.

Brod, R. I., & Devens, M. S. (1985). Foreign language enrollments in U.S. institutions of higher education-Fall 1983. *ADFL Bulletin, 2*, 57-63.

Foreign language enrollments in public secondary schools - Fall 1982. (1984). *Foreign Language Annals, 17*, 611-623.

Galloway, V. (1986). From defining to developing proficiency: A Look at the decisions. In H. Byrnes & M. Canale (Eds.), *Defining and developing proficiency: Guidelines, implementations, and concepts* (pp. 25-73). Lincolnwood, IL: National Textbook.

Hiple, D. V. (1986). A progress report on the ACTFL proficiency guidelines, 1982 - 1986. In H. Byrnes & M. Canale (Eds.), *Defining and developing proficiency: Guidelines, implementations, and concepts* (pp. 5-24). Lincolnwood, IL: National Textbook.

Interagency Language Roundtable language skill level descriptions. (1985). Arlington, VA: US Government Interagency Language Roundtable.

Kahane, H. (1986). A typology of the prestige language. *Language, 62*, 495-507.

Liskin-Gasparro, J. E. (1984). The ACTFL proficiency guidelines: A historical perspective. In T. V. Higgs (Ed.), *Teaching for proficiency: The organizing principle* (pp. 11-42). Lincolnwood, IL: National Textbook.

Oral language proficiency testing at the Foreign Service Institute. (1986). Washington, DC: Department of State, Foreign Service Institute School of Language Studies.

Wilds, C. P. (1975). The oral interview test. In R. L. Jones & B. Spolsky (Eds.), *Testing language proficiency* (pp. 29-44). Arlington, VA: Center for Applied Linguistics.

APPLICATION OF THE ILR-ACTFL TEST AND GUIDELINES TO THE LESS COMMONLY TAUGHT LANGUAGES

Ron Walton
University of Maryland and National Foreign Language Center

The application of the ILR-ACTFL Guidelines to the less commonly taught languages (LCTL) has been an interesting, and quite challenging, undertaking. To my knowledge, the earliest meeting on this topic sponsored by ACTFL included language teachers and scholars of Chinese, Japanese, and Russian. What follows is a brief overview of some of the issues that came up as this group entered into the task of preparing proficiency guidelines for these languages.

To be honest, some members of this working group came to the task with more than a little skepticism. This is perhaps due to a rather common perception among those involved with the less commonly taught languages that these languages are so difficult, and/or so unique, that they somehow inherently defy any sort of standardized evaluation procedure. Nevertheless, the group as a whole felt that some sort of standard evaluation metric was quite necessary, and in fact, long overdue.

The group's initial brush with the ACTFL-ILR Generic Guidelines was not particularly controversial. There was agreement, however, that the development of language specific guidelines would most likely prompt a more serious look at the generic guidelines as work proceeded. There was certainly never any doubt of the need for language specific guidelines, however. Unlike DLI or FSI which test in-house and which can train their testers using just generic guidelines, the very variety and number of academic programs, as well the enormous number of potential examiners, is such that generic guidelines are simply too broad and vague for purposes of consistent, reliable, standardized evaluation of language competency. Moreover, it is often the case that native and non-native teachers of the LCTL's have different ideas and approaches as regards testing, and this too argued for a rather explicit set of language specific guidelines agreeable to both teaching populations.

As an example of procedure in working on the language specific guidelines, some members of the Chinese team used the German language-specific guidelines as a rough model. It was during this step that the first problems with the generic guidelines began to emerge. These same problems were encountered by the Japanese and Russian teams, and eventually, as is well known, the generic guidelines were revised to their present state. The result of this latter endeavor was to eliminate a slight Indo-European language bias, mainly with respect to linguistic features such as "tense" (as opposed, for example, to aspect in Chinese and Russian), and to features of writing systems relevant to the reading and writing guidelines. It is, moreover, quite possible that the generic guidelines may someday need even more refinement as they are applied to languages from other language families.

Much more problematic was the attempt to develop language specific guidelines for the LCTL's at the lower proficiency levels, particularly the Novice and Intermediate Levels. Here, the working groups were confronted with a genuine puzzle: ACTFL was attempting to provide rating gradations at the lower end of the proficiency scale that were much more detailed than the ILR Levels of 0 and 1. Yet, because of the time required to get to a ratable level in the LCTL's, the LCTL guidelines meshed very poorly with the commonly taught language guidelines. For example, the performance descriptions for a Novice-Mid in Spanish seemed to correspond more to, say, an Intermediate-Mid in Chinese. Of course, the LCTL teams were quite aware that at an abstract level, there can be no differences on a proficiency scale based on the features of a particular language. An Intermediate in French is no different than an Intermediate in Japanese. Certainly, the government language teaching agencies have applied one and the same scale to all languages and the proficiency rating system would not even be a system if this were not the case.

The problem is largely one of the contact time necessary to reach a certain stage, at least as far as formal language study programs are concerned. It simply takes longer for learners in some of the LCTL's to reach the same level as learners in the more commonly taught languages. Of particular concern to the LCTL guideline group was that the time lag was so great that students of the LCTL's in academic programs would be depressingly low on the ACTFL-ILR scale, even after considerable time (compared to the more commonly taught languages) spent in studying the language. Thus, a Spanish student in a university program might test out as an Advanced-Low, but the student of Japanese, having studied the same amount of time, might reach only the Intermediate Level. This situation is, of course, the reality, but there was a feeling that neither teachers nor students would welcome being confronted with it via a standardized evaluation metric.

Part of the contact-time problem, for some of the LCTL's, is related to the difficulty of the writing systems. Since most formal programs attempt to teach all four skills simultaneously, the extra time required in dealing with complex orthographies not only makes movement in reading and writing lag behind languages with simpler orthographies, but there is less time to devote to oral-aural skills and development of these skills may be retarded as well.

Different solutions were proposed to overcome these problems. The view of some, including myself, was and is that the lower level guidelines for the LCTL's should be dropped for learners in formal study programs, but retained for the testing of "acquirers," those have picked up the language informally. This would mean that learners in formal programs are simply not given proficiency tests until there has been enough exposure and work in the language to make a decent showing on the scale. In any case, the lower level guidelines were retained, but with some rather interesting adjustments.

These adjustments center on the reading and writing guidelines. A distinction was made between "authentic" and "specially prepared" reading material for instance. Over this distinction, was laid another between "reading" and "decoding" (and/or the use of a dictionary). This allows giving a rating to a subject who can "decode" (but not "read") an "authentic" passage, but can "read" (not decode) a "specially prepared" passage. Over these two distinctions, yet another was added, at least for Chinese, between reading/writing material that more closely resembles colloquial or everyday speech, and material that is more characteristic of genuine writing style. That is, reading/writing material was put on a continuum in, for example, Chinese, with spoken style on one end, and very literary style, having elements of Classical Chinese, on the other end. The thinking here was that in formal learning programs which teach speaking, listening, and modern spoken grammatical patterns and lexicon, students could carry this knowledge over to reading and writing "speech-like" material. However, as learners move up the scale, they will increasingly encounter structure and lexicon more characteristic of literary style which can differ dramatically, in some languages, from speech. While I pushed these distinctions and put considerable effort into developing them for the Chinese reading and writing guidelines, I realize that they defy some principles of proficiency evaluation, and, in fact, that they start to make the guidelines look rather achievement oriented. Nevertheless, these and similar devices make it at least possible for LCTL learners in some languages to get on the scale in a way comparable to those studying the more commonly taught languages.

Yet another problem concerning the written medium skills were low level descriptions obviously geared toward measuring the abilities of acquirers, as opposed to learners in formal instructional settings. The generic guidelines and the language specific guidelines for the commonly taught languages often seem to imply that some abilities come from exposure, in a somewhat natural way. For example, low level reading and writing descriptions seem to suggest that a person would be able to read signs, write his or her name, copy down the names of streets and the like. This may be true for French, for example, where there are many cognates and where an alphabetic writing system is already familiar to the person. However, there is no guarantee that exposure to Chinese characters will lead to even a primitive ability to read them. Signs, for example, quite often even make use of a more literary style of language that is somewhat telegraphic (this saves space). Letters in Chinese likewise often follow a semi-literary template. Moreover, to write or copy down a Chinese character requires a knowledge of stroke types and stroke order which is anything but "natural" to a native English speaker. In short, some facets of reading and writing in some LCTL's simply seem beyond informal acquisition and require instead formal training. Thus, trying to write the low-level guidelines geared toward "picking up" certain abilities from exposure proved to be rather unsatisfactory, at least for some LCTL's.

A final problem to be mentioned concerns the testing and evaluation of socio-cultural features in the guidelines for oral-aural skills as manifested in the oral interview. The ACTFL/ILR guidelines for culture were eventually abandoned, largely because, I believe, we simply lack the theoretical knowledge to go about such an endeavor. On the other hand, one could make the case that there should not be a separate set of culture guidelines, but rather that socio-cultural abilities in face-to-face communication should be spelled out in the speaking-listening guidelines and evaluated during the oral interview. In any case, it seems likely that the LCTL's will again present many special problems, perhaps only of degree, compared to the more commonly taught languages : just as the LCTL languages can differ greatly with respect to linguistic features from the commonly taught languages, they can differ even more significantly with regard to socio-cultural interpersonal verbal behavior. It could well turn out that what needs to be evaluated in the socio-cultural verbal interaction dimension will be much more elaborate and complex in the LCTL's than one might assume from looking first at the Indo-European languages. So much so, that I would suggest that ACTFL or anyone else working on this dimension begin, this time, with the LCTL's and work into the Indo-European languages rather than the other way around, a procedure that hindsight reveals would have made more sense with all of the ACTFL/ILR Guidelines.

One particular problem that has been debated for many years in the LCTL teaching community must be considered if the cultural dimension is to be further developed in the guidelines. The proficiency guidelines are based on the non-native's proximity in language competency, and presumably in cultural competency, to an educated native speaker. Yet, for some LCTL cultures, the case could be made that the closer the non-native resembles the native speaker, the greater the possibility for a problematic communication interface.

This issue may not come up for the Indo-European languages because a non-native speaker may stand some chance of not being perceived as a obvious foreigner. In some LCTL cultures, however, the foreigner is always and obviously a foreigner, regardless of linguistic ability. In this case, it is possible that native speakers have a separate and well-defined, though quite implicit, code or registry for the appropriate socio-cultural verbal behavior of foreigners, as foreigners. If the linguistic competence of a foreigner far outstrips his or her cultural competence, and this does happen, the situation can be particularly troublesome, such that increasing linguistic competence without the accompanying cultural skills leads to diminishing returns in communication interactions.

Therefore, in future work on the culture dimension of the Proficiency Guidelines, it may become necessary to approach the LCTL's in an unorthodox manner compared to the commonly taught languages. It could turn out, for example, that a more realistic procedure is to peg the base language speaker's abilities to an "ideal foreigner" rather than to an educated native speaker of the target language. The idea would be that as one moves up the scale, one is not becoming more like a native speaker, but one is becoming more adept at making a native speaker comfortable in interacting with an obvious foreigner.

To sum up, we have developed some guidelines for the LCTL's and have come across enough interesting problems in the process to make us think that this is just the beginning of an on-going endeavor. I believe that the Guidelines can certainly be applied, with proper training, but I suspect that the lower level guidelines will continue to be vexatious to many. The "tricks" that some of us pulled off in the lower level descriptions can't be all that satisfying to proficiency purists and some of us would rather have skirted these problems by dropping the low-level descriptions completely, except for acquirers.

To the extent that my thinking represents that of at least some colleagues in the LCTL field, the formal study of some LCTL's just seems so achievement-bound that proficiency testing at the lower levels simply isn't worth the effort. We in the LCTL's often seem to be confronted with what I term "systemic thresholds." We do want meaning to take precedence over form, and we do begin "communicating" in classrooms from the first day, but we deal with certain linguistic systems or subsystems that seem to require early and intensive attention to form as such. Our feeling, I think, is that there are minimal thresholds in the instructional process for such systems as pronunciation, orthography, romanization/transcription, stroke types, stroke order, and character components for Chinese characters, and the like, and that these should be somehow "under control" almost as a prelude to serious work on meaningful communication. Because many of the concepts that underlie these "systemic thresholds" are quite alien to Indo-European speakers, more time is required for communication readiness, and thus the less relevant become the low level proficiency guidelines which seem to assume urgent and immediate needs to communicate in the real world.

In any case, we've made a start and I'm sure all of us are interested in seeing these guidelines used and in seeing them refined. It's hard to imagine, from my personal viewpoint, that the LCTL language teaching community would allow the various problems that I've touched on to prevent us from recognizing the benefits of a standardized metric, though we will, and should, continue to debate the nature of this metric.

BENNETT-BIERSTEKER MODEL FOR PROFICIENCY PROFILING IN AFRICAN LANGUAGES

Sally Sieloff Magnan
for
Patrick Bennett
University of Wisconsin-Madison

As a past consultant to a proficiency project for developing a testing instrument for African languages, I have been asked to summarize briefly an "alternate" or "additional" version of the ACTFL/ETS/ILR Oral Proficiency Interview (OPI) suggested by Patrick Bennett of the University of Wisconsin-Madison and Ann Biersteker of Yale.

The Bennett-Biersteker model for proficiency testing in African languages is currently under revision, following an intensive three-week workshop in May 1986. As currently planned the model includes an OPI, based on the ACTFL model, and role-play in cultural scenes. Bennett and Biersteker call their model "proficiency profiling." Based on speech act theory, their model aims to provide more precise information than is obtainable through the OPI, in order to be more diagnostic for students taking courses. Profiling is to be used in situations where finer diagnosis is needed and it is to accompany a global ACTFL rating.

The Bennett and Biersteker model evaluates language in two channels (aural/oral and reading/writing) and hence four modalities, and also three aspects of language not measured separately by the ACTFL OPI. Bennett and Biersteker call these three areas socio/cultural, lexical/traditional, and emotive/pragmatic. They are defined as follows:

> - Socio/cultural --cultural specific units of activity involving patterned behavior, including greetings, introductions, common transactions, and formalized activities such as games, interviews, and social functions;
> - lexical/traditional --the range of cultural vocabulary, use of idioms and cultural allusions, the stock of expressions, references, knowledge that pertains to literary, historical, political, and cultural heritage;
> - emotive/pragmatic --control of the force of communication, of how we send, receive, and share signals about the purpose of language through irony, sympathy, sincerity, deference, lying, etc.

In addition, each area (the four modalities and three aspects just discussed) is analyzed as three types of linguistic behavior: one-way (monologue), interactive (role-play), and abstract (knowledge of process) to give a complex profile of 36 bits of information or 36 separate evaluations, each based on a three-point scale that measures degree of competence or control.

In their workshop at the University of Wisconsin-Madison in 1986 on testing in Arabic and Swahili, Bennett and Biersteker received preliminary feedback on their model that identified problems in both the interview structure (the standard ACTFL procedure) and in the rating (a combination of ACTFL global rating and profile rating). These are problems that will have to be faced by all models that are comprehensive in theory and detailed in evaluation.

With regard to the interview, the ACTFL OPI did not provide enough data -- certainly not in reading and writing, but also not in listening, in listening-speaking interaction, and in the cultural-based notions of lexical/traditional and socio/cultural. The problems with reading, writing, and even listening aside, the Bennett and Biersteker model requires an interview expanded beyond the ACTFL model to include many more interactive exchanges in pseudo-authentic cultural contexts.

With regard to rating, the Bennett and Biersteker preliminary model was difficult to use in that it required testers to make up to 36 distinct evaluation decisions, an apparent problem for rater reliability. Furthermore, the three-point scale provided only a narrow range of measurement (up to ACTFL Advanced Level it seemed). An even more detailed system, perhaps a four-point, scale would be needed to make distinctions equivalent to ACTFL Superior Level.

Since the May 1986 workshop, Bennett and Biersteker have continued to work on their model. In May 1987 the African Studies Program will host a second workshop to discuss proficiency testing for African languages in general. Bennett and Biersteker expect that this workshop will offer a forum for a critical look at their proposed model, particularly with regard to suggested changes in the model itself and in the rhetoric of the rating descriptors. Following this workshop, the African Studies Program hopes to publish a report that will summarize the past two-year effort in developing a suitable proficiency test for African languages.

ARABIC-SPECIFIC ORAL PROFICIENCY GUIDELINES

Ernest N. M^c^Carus
University of Michigan

Four proposals are made in reference to Arabic-specific oral proficiency guidelines: 1. After a brief description of Arabic diglossia, that the guidelines be designed for Arabic as a whole, permitting and expecting the testee to respond in either Modern Standard Arabic or a dialect as appropriate; 2. that language-specific guidelines not be written in terms of specified language features but in terms of grammatical accuracy, communicative proficiency, acceptability to native speaker, and diglossic appropriateness; 3. that a critique of language errors be provided the testee after each interview; and 4. that a multi-component score be provided in addition to a global one.

I should like to present four comments on the Oral Proficiency Interview with particular reference to Arabic. But first I would like to describe briefly my own experience with the FSI oral interview test.

CASA, the Center for Arabic Study Abroad, provides a summer and an academic year of intensive, full-time Arabic language training for advanced-level American students, carried out at the American University in Cairo. In the early 1980's when I was director we brought Dr. Margaret Omar from the FSI School of Arabic in Tunis to train our staff, all Egyptians, in the FSI oral interview technique. She made several trips to Cairo over the period of two years or so to instruct us, at the end of which she felt that one instructor at least was certifiable by FSI standards. By the end of her training our individual ratings of testees agreed quite closely with hers. We found it useful to introduce decimals to provide a necessary degree of gradation in the scores, such as 1.8, 2.5 or 3.3. Incidentally, after one year of study of Egyptian colloquial and Modern Standard Arabic students generally raised their level one whole integer, e.g., 1.8 to 2.8 or 3. Testees could respond in either the colloquial or the literary language as they wished. We noted not only grammatical accuracy and communicative proficiency but also pragmatic and sociolinguistic factors--intonation, kinesics, etc. After each interview the students' strengths and weaknesses were discussed with them; this critique of the testees' weaknesses and strengths was considered by both teachers and students as an important by-product of the test.

1. Diglossia in Arabic

In his paper David Hiple referred to the problem of diglossia in connection with setting up language-specific oral proficiency guidelines for Arabic. Diglossia is an enormously complicating factor in the learning and teaching of Arabic; it will be useful here to make a few remarks about it.

Arab children learn at home the dialects peculiar to their family's communities. These vary according to geographical location, socio-economic class and education, religion, sex, etc. At school they learn the literary language, usually called Modern Standard Arabic (MSA).[1] Dialectal or "colloquial" Arabic is used for day-to-day oral communication, while MSA is used in formal situations: lectures in the university, newscasts on the air, discussing technical topics for which colloquial Arabic lacks the terminology, or for any kind of writing. The Arab World today is roughly like Italy before Dante: vernacular Latin was used then for daily basic needs but Literary Latin was used for formal occasions or for writing. Colloquial Arabic is used between Arabs who speak the same dialect or variety of Arabic in most situations; they will switch to MSA to discuss technical topics such as political or economic policy, religious matters, etc. to the extent that their dialect cannot handle them, or in formal situations such as delivering a speech before an audience. On the other hand MSA is used between speakers of different dialects, especially those that are geographically far removed from each other and so are not so easily understood, at international conferences, etc. In any case MSA is always used for writing, with a colloquial dialect used only occasionally in drama or for dialogue in fiction. To borrow from computer terminology, colloquial is the default form for speaking, whereas, in linguistic terms, MSA is marked for [+formal]. These roles are reversed when it comes to writing, where MSA is the default style and the colloquial is now marked for [+dialect].

This creates a major problem in proficiency testing. Programs of instruction in the United States typically teach MSA; only the larger institutions include any dialect training. Thus, a graduate of the Middlebury summer intensive program in Arabic might be an Intermediate High in learned discussions but an Intermediate Low or even Novice High on many basic survival needs, since MSA is not normally used for these. Various solutions have been considered: (a) ignore the dialects and write the guidelines for MSA only; (b) write guidelines for MSA and for each of the major colloquials; and (c) choose a major dialect--say, Egyptian, since most Americans study Egyptian--and write a set of guidelines for it as well as for MSA.[2]

My proposal is that we treat Arabic as a single living language and deal with it as we do with other living languages, with a single set of guidelines that cover all real-life situations. The testee may respond with either literary or colloquial Arabic, according to what he or she knows and according to which one is appropriate for a particular occasion, at any stage of testing. The rater then factors the matter of the diglossic appropriateness of the response into the student's score, taking into account the use of the colloquial where native speakers use colloquial and of MSA where MSA is normally used. This provides ample leeway for the individual who is fully proficient (which by definition would mean proficient in both MSA and a dialect) to perform in all phases of the test and at the same time permits testing of those who know only one or the other form of Arabic on a more limited scale. Those who know a dialect only could easily be rated, since they will in all probability cluster at the Novice and Intermediate levels, but those proficient in MSA may well present a rater's dilemma: they may rate an Advanced Low on the formal side but only a Novice High on some basic skills--would they still merit an Advanced Low?

2. On defining proficiency levels

I would like now to comment on the language-specific proficiency guidelines. I propose here the goal of defining each proficiency level or sub-level in terms of (a) functional skills, such as "able to make an appointment over the telephone"; (b) acceptability to native speaker, such as "native speaker often fails to understand what is said"; and (c) degree of linguistic accuracy, for example "has mastered all the non-English phonemic contrasts" or "has mastered the more common inflections of the verb", as diagnostics of the various levels. I am uneasy with the use of specific grammatical or lexical features as representative of particular levels. Even though they are described as suggestive and not prescriptive the irresistible tendency will be for the individual to look for these features and base ratings on them. For example, in a particular set of guidelines for Arabic the conditional sentence--"If he comes I'll tell you"--is mentioned at the Advanced and Superior Levels; since most textbooks devote special attention to it this is a logical thing to do. It is a complex construction for English speakers, so some teachers or programs of instruction delay its presentation or underteach it. Others, again because of its difficulty, might overteach it successfully but to the exclusion of other constructions. If the conditional sentence is taken as diagnostic for either of these two groups of students the ratings will be skewed.

Is it not possible first to define and illustrate, before presenting the guidelines themselves, the grammatical categories of the language and then within the guidelines talk in terms of degree of mastery? For example, phonology for Arabic includes in addition to stress and intonation linear phonemes with the special problems of "emphasis" (pharyngealization), velar and pharyngeal fricatives, etc; derivational morphology includes participle and verbal noun formation as well as the famous extensive derived verb system; and style includes saying instead of "five years after the war", which is understandable but not idiomatic, the equivalent of "after the war by five years". Then at the appropriate levels say such things as "complete mastery of phonology"; "nearly complete mastery of inflectional morphology;" or "heavy English interference in idiomatic expression." There is so much variation in order of presentation of grammatical features from one class to another that to suggest or assume a particular order of presentation is highly arbitrary and potentially misleading. On the other hand the Chinese-, Russian- and Arabic-specific Guidelines that I have seen all provide ample discrete linguistic items typical or suggestive of each level, which implies that the approach I am suggesting here is impracticable. But has it been adequately tried? Perhaps after extensive trial-and-error experience we will reach the stage where it will indeed be possible to construct such a set of guidelines.

3. The OPI as a diagnostic test

I was happy to note from some of the presentations yesterday that others use the Oral Proficiency Interview to provide feedback to the students on their progress--in particular, those areas that need special attention and effort--in learning the language. While the OPI is not specifically designed for that purpose such diagnostic power is nevertheless a valuable by-product that should be recognized for its worth and exploited fully.

4. Global and multi-component scores

It follows that I am in favor of a multi-valent or multi-component score in addition to the global one for the additional information that it provides to the learner and to his or her teacher. There should be a rating for grammatical accuracy (and in some cases a further breakdown into phonology, morphology, syntax and lexicon); communicative proficiency (in terms of levels of skills that can be performed); acceptability (intelligibility to native speakers, pragmatics, and kinesics); and appropriateness: competence in and appropriateness of choice between MSA and colloquial Arabic (functioning in a diglossic environment).

Notes

[1] MSA has also been referred to as "Literary", "Contemporary" and "Classical" Arabic.

[2] See the discussions in Parkinson (1985) and Allen (1985).

References

Allen, R. (1985). Arabic proficiency guidelines. *Al-'Arabiyya, 18*, 45-70.

Parkinson, D. B. (1985). Proficiency to do what? Developing oral proficiency in students of modern standard Arabic. *Al-'Arabiyya, 18*, 11-44.

THE CERTIFICATE OF DUTCH AS A FOREIGN LANGUAGE

Jos Nivette
Free University of Brussels (VUB)

Dutch can be geolinguistically defined as the usual language in the Netherlands (with 14 million inhabitants) and in the northern part of Belgium (6 million inhabitants), albeit with two exceptions: in the north, Frisian is spoken as the official language, and in the south, in Brussels, there is a majority of French speakers. Outside Europe Dutch is spoken in the Antilles and traces of the language remain in Indonesia. Finally, Dutch also gave rise to the only Germanic language which is exclusively spoken outside Europe, namely Afrikaans, which 50 years ago replaced Dutch as one of the official languages of the Republic of South Africa.

Since 1946 the cultural integration of the Netherlands and Belgium has been actively promoted by various official and private initiatives. The youngest initiative in that respect has been the creation in 1980 of the Dutch Language Union by the two governments. The Dutch Language Union is founded on the unity of the language spoken and written in the Netherlands and in the Dutch-speaking part of Belgium.

Therefore, the aims of the Language Union are namely cooperation for the development of the language, its prestige, its influence abroad, its dissemination. Work concentrates on a common terminology for legislation and official publications, and on the practical organization of courses in the Dutch language, literature, and culture abroad.

The fact is that the Dutch language is part of the curriculum in about 200 universities abroad and in 30 different countries. In Europe alone there are about a hundred universities concerned, of which 25 in Germany, 13 in the UK, 10 in Italy, 6 in Poland. Furthermore almost all European countries are represented in the list of Dutch chairs abroad. (Among these countries Bulgaria, Denmark, Finland, Hungry, Norway, Austria, Portugal, Rumania, Spain, Czechoslovakia, Sweden, Switzerland.)

On the part of those concerned with the teaching of Dutch abroad there has in the past been considerable demand for an officially recognised Dutch exam for non-native Dutch speakers. This proposal was supported ten years ago by the Belgian and Netherlandic governments and as a "Certificate of Dutch as a Foreign Language" was officially introduced in 1977. The Language Union Treaty in 1980 confirmed and strengthened the official status of the certificate: the Treaty states that "both governments agree upon common norms of testing for the common awarding of the certificate".

The "Certificate of Dutch as a Foreign Language" is a unit-credit certificate scheme which enables learners of Dutch to acquire separate certificates on different levels. Examinations are for elementary knowledge, basic knowledge and advanced knowledge. At each level separate examinations are provided for the four skills: listening, reading, writing and speaking. Credits may be acquired for each of the four skills, and candidates who have acquired the four credits belonging to one level receive a unit certificate for that level. The unit-credit system is not absolute since candidates may also take the tests for the "four skills" at the same time, or they may directly go for the basic level without having acquired the elementary level.

The examinations of the Certificate of Dutch are long-distance examinations in that they are prepared and corrected by a central examiner's board and administred by local examiners. The scientific and organizational aspects of the project have been entrusted to a Dutch-Belgian team, based in the Université Catholique de Louvain in Louvain-la-Neuve. From Louvain-la-Neuve the exam papers are sent to the examiners. As a rule they are professors of Dutch abroad.

The examiners all receive the same assignments and send the candidates' copies back to Louvain-la-Neuve where the decision as to the awarding of certificates is taken.

The form of the certificate examinations however makes special demands upon the test format. The fact that examinations are centrally organised and corrected has certain disadvantages with respect to the administration. Speaking proficiency e.g. cannot be tested by means of an oral interview test as we cannot always make use of trained testers and as it is not economically feasible to send a testing team to the various institutes. Special care has to be taken that the test can easily be administered by the local teachers and that yet the test conditions remain the same for all testees, no matter where they take the test.

In this way, an international unity for the rewarding of the certificate can be reached, and it does not matter whether the applicant takes the exam in Paris, Berlin or in Alabama.

Notes

[1] Some information in this discussion is taken from L. Beheydt, *Testing for the "Certificaat Nederlands als vreemde taal"*, UCL, Louvain-la-Neuve (mimeogr.)

[2] More detailed information about the tests can be obtained at the following address:

Certificaat Nederlands als vreemde taal
U.C.L.
Collège Erasme (C 387)
Place Blaise Pascal, 1
B - 1348 Louvain-la-Neuve
België

COMMUNICATIVE NEEDS: THE CASE OF IRISH[1]

David Singleton
Trinity College, Dublin

Communicative approaches to second and foreign language teaching and testing typically begin from an assessment of learners' communicative needs. This paper examines the concept of "communicative need" as it relates to the learning of Irish, especially in the light of the limited number of remaining native speakers of the language and the fact that they are all bilingual. The paper argues that "need" in this context must be broadly interpreted, so as to include expectations and interests, and that "communicative" must be seen as covering "communication" with "Gaelic culture".

A defining characteristic of the communicative approach to language teaching, at least as it is generally understood in Europe, is that it is learner-centered. That is to say, it endeavours to set learning objectives not in terms of an abstract body of knowledge (e.g. the target language system), but in terms of the learner's own purposes, experience, style and aptitude. For this reason, the "classic" communicative syllabus starts from an assessment of the particular communicative activities a given group of learners with a given set of purposes will need to be able to cope with in order to fulfil those purposes. On the basis of this assessment of communicative needs, general aims and specific behavioural objectives are elaborated, and the latter are then broken down into socio-semantic categories (notions and functions), the formal exponents of which constitute the lexico-grammatical specification of the syllabus. As for communicative tests, whether formative or summative, their content is determined by the syllabus, and can thus also trace its origin back to learner needs.

For those of us formed in the Council of Europe school, then, the very foundation of the whole communicative edifice is the analysis of learners' communicative needs. Some unease has been expressed, here and elsewhere, about the fact that the ACTFL Guidelines, which in other respects seem sympathetic to the communicative perspective, do not draw on any such analysis. I have to say that I share that unease. However, I should also like to dissent from any suggestion that assessing learners' needs is in all cases a straightforward affair. It is worth remembering that the Council of Europe model of syllabus definition was initially constructed with adults in mind, and indeed primarily adults who could identify immediate and concrete circumstances in which it would be absolutely necessary for them to be able to function through their target language (see, e.g., Richterich, 1973). Discovering and taxonomizing communicative needs in such instances does not seem particularly problematic. However, how is one to proceed when language learners are not motivated by functional imperatives? The difficulties that can arise are well demonstrated by the case of Irish.

Irish, as is well known, is a language in deep decline. According to one recent estimate (Ó Murchú, 1984) Irish is the first language of no more than about 100,000 people living in the Republic of Ireland, that is, about 3% of the total population. Most of these native speakers are domiciled in thinly populated rural districts of Counties Donegal, Mayo, Galway, Kerry, Cork and Meath, such Irish-speaking pockets being collectively known as the "Gaeltacht". This term is actually potentially misleading, since it might be taken to imply that the inhabitants of the areas in question speak only Irish, whereas, in fact, all are bilingual, that is, also have a native or near-native command of English.

Accordingly, an English-speaker going to the Gaeltacht has no real need to use Irish in order to "survive" or even to socialize. While Anglophone visitors to France, Germany, Italy, Spain, etc. are aware (or quickly become aware) that as soon as they stray beyond the tourist traps they will be seriously constrained if they have only English at their disposal, visitors to the Gaeltacht know full well that English will see them through in all circumstances. Indeed, one problem that learners of Irish face when visiting the Gaeltacht with linguistic immersion in mind is that, if ever they evince signs of communicative difficulty in Irish, their native speaker interlocutors, whether from impatience or kindness, tend to switch to English.

Outside the Gaeltacht Irish is seen but rarely heard. Bilingual street-signs, public notices and official documents are the rule, but more ephemeral communicative acts in Irish are very much the exception. There is in the non-Gaeltacht population a tiny percentage of native speakers - mostly Gaeltacht "émigrés" - who continue to use Irish, as well as a scattering of non-native speaker enthusiasts who speak Irish wherever possible. There is also a small number of institutions and organizations which operate partly or wholly through the medium of Irish[2]. Most of these are state-funded, and almost all perceive themselves as playing some part in Irish language revival, which since the foundation of the Irish State has had government support.

Otherwise, spoken Irish outside the Gaeltacht is confined to residual formulas, ritual lip-service, and various kinds of artistic activity. To the first category belong stock greetings, toasts and leavetakings in Irish, to the second the incorporation of a paragraph or two of Irish into certain kinds of speeches, and to the third song-recitals, poetry-readings and plays in Irish.

It is evident in the light of the foregoing that the teaching of Irish as a second language in Ireland has little to do with providing for English-speakers' communicative needs in the Gaeltacht or elsewhere, and a very great deal to do with sociolinguistic engineering, that is to say language maintenance and revival. Indeed, the Irish program in the schools probably constitutes the last line of defense for the language. It must immediately be added that such efforts to maintain the language have, according to recent surveys (see, e.g., CLAR, 1975; Ó Riagáin & O Gliasáin, 1984), at least the passive support of the majority of the Irish population.

However, other studies have shown that, when asked about the practical uses to which Irish might be put, past and prospective learners of Irish are nonplussed. For example, a question concerning the practical benefits of Irish in daily life on a questionnaire administered to a group of subjects about to embark on a broadcast multimedia Irish course was simply ignored by 38.6% of the group, and a large number of the responses it did elicit were extremely vague ("future use throughout daily life", etc.). The remaining responses were divided between the uninformatively general (e.g., "for encouraging others to speak Irish"), and the educationally or culturally oriented (e.g., "for helping the children with their homework", "for reading", "for understanding songs")

(Devitt, Little, Ó Conchúir & Singleton, 1982-3, p. 44).

Likewise, in a survey of the language experience, attitudes and needs of undergraduate and postgraduate students at Trinity College, Dublin, Little, Singleton & Silvius (1984, pp. 185-187) found that, faced with a question about circumstances of possible use of Irish, 50% of undergraduates who knew Irish but said they would like to know it better failed to give any reply, and 70% of postgraduates in this category did not respond. Of the responses that were elicited from such students, many referred to cultural and educational pursuits. With regard to students who knew no Irish but said they would like to learn some, most did actually answer the question, and references to cultural and educational circumstances again loomed large. Similar findings have more recently emerged from a small-scale study of school students in their early to mid teens (Fox, 1986).

Such results underline the difficulty of identifying needs -- at least in the "classic" sense -- for learners of Irish. The consequence of this difficulty for those endeavouring to apply communicative principles to the teaching of Irish is a choice between abandoning learner-centeredness and redefining "needs". Given that learner-centeredness is such a cornerstone of the communicative approach, the former course would be difficult to justify. The latter, when one looks at it closely, does, however, offer possibilities.

For one thing, one can consider the needs the learner may have in the context of learning. Irish is usually taught at least partly through the medium of Irish; sometimes other subjects are also taught through Irish. It therefore seems reasonable to claim that learners of Irish genuinely *need* access to classroom discourse in Irish. This may be a need imposed by the system, but so, after all, are many of the needs one confronts outside the school walls! Broadening the focus somewhat, one might consider learners' expectations and general interests. In relation to expectations, even if it is never strictly necessary to use Irish in order to "survive" or socialize the learner of Irish will undoubtedly *expect*, nevertheless, that a program in the language will equip him/her to engage in basic transactions and interactions in Irish. As for the matter of general interests, within any group of learners a common pool of such interests can fairly readily be established. Some will relate closely to the Irish language (notably the widespread interest in "Gaelic culture"); others can at least be treated through the medium of Irish. Expectations and interests are not *stricto sensu* needs, but they do reflect learners' perceptions and preoccupations, and are therefore, in a learner-centered approach, perfectly appropriate candidates for inclusion as content-determining criteria.

The concrete attempts that have been made to bring communicative principles to bear on programs in Irish (see, e.g., Comhar Na Múinteoirí Gaeilge, 1982; Norton, 1984; Little, Ó Murchú & Singleton, 1985a, 1985b) have adopted the above line, realizing learner-centeredness by reference to students within-school needs, learning expectations and general interests.

Moreover, the same kind of approach has characterized certain of the syllabuses devised in Ireland for foreign languages (see the French and German Intermediate Certificate syllabuses, Department of Education, 1985-6). The reason for this is simply that, since only about 10% of "general" learners of foreign languages in Ireland actually visit countries where those languages are spoken, 90% of learners find themselves in exactly the same position in relation to those languages as in relation to Irish, that is, without genuine communicative needs beyond the confines of the classroom. In other words, "general" foreign language learning also poses problems for "classic" needs analysis and also calls for a broad interpretation of "needs".

To return to the widely attested educational and cultural interest of learners of Irish, these may induce one to examine the notion of "communication". Many learners of Irish evidently wish to "communicate" not so much in a face-to-face mode with individual native speakers as in more distant modes with what they conceive of as "Gaelic culture". Can such "communication" be licitly encompassed within the communicative approach? A moment's reflection will persuade us that it can. Communication with a culture cannot but imply communication with individual exponents of that culture - poets, dramatists, singers, etc. Furthermore, it has often enough been emphasized that, as envisaged by the communicative approach, "communication" covers reception as well as production, writing as well as speech and discourse which is planned as well as unplanned, non-reciprocal as well as reciprocal. Accordingly, communicative principles offer absolutely no obstacle to providing for learners of Irish to gain access to what they perceive as their cultural heritage. Again, this discussion of the case of Irish can be seen as having relevance for the more widely taught languages such as French and German. Some learners of these languages are principally interested in the art they mediate. What the above consideration of the "Gaelic culture" question clearly shows is that such interest can readily be accommodated within the communicative approach.

To conclude, the particular sociolinguistic circumstances of Irish pose problems for the notion of basing second language programs in that language on communicative needs analyses of the "classic" kind. Similar problems may be presumed to arise in relation to all the Celtic languages and indeed to any minority language most or all of whose speakers also have a native-like competence in a very widely spoken language. Such problems are not, however, insuperable, given a suitably elastic understanding of "needs" and a fully rounded definition of "communicative". Nor does this broad interpretation of "communicative needs" find application only in respect of minority languages; it has relevance to curriculum issues which touch the more widely spoken and more widely taught languages also.

Notes

[1] I should like to thank David Little for his very helpful comments on the first draft of this paper.

[2] Included in this category is the Irish national broadcasting service, Radio Telfis Eireann, which provides an all-Irish radio channel (Radio na Gaeltachta), as well as occasional programs in Irish on its three mainly English-medium radio channels and two mainly English-medium television channels. Radio na Gaeltachta, as its name implies, is aimed principally at the Gaeltacht. Irish language enthusiasts are continually complaining about the paucity of Irish language programs on the other radio channels and on television.

References

CLAR (Committee on Language Attitudes Research) (1975). *Report*. Dublin: Government Publications.

Comhar na Múinteoirí Gaeilge. (1982). *Mise agus tuse*. Dublin: Author.

Department of Education. (1985-6). *Rialacha agus clar do leith meanscoileanna*. Dublin: Government Publications.

Devitt, S. M., Little, D. G., Ó Conchúir, S. P., & Singleton, D. M. (1982-3). *Learning Irish with Anois is Arís* (CLCS Occasional Paper). Dublin: Trinity College, Centre for Language and Communication Studies in association with Radio Telefis Eireann.

Fox, G. (1986). *Preliminary explorations of some post-primary pupils' attitudes to Irish.* Unpublished graduate diploma dissertation, Trinity College, Dublin.

Little, D. G., Singleton, D. M., & Silvius, W. F. M. (1984). *Learning second languages in Ireland: Experience, attitudes and needs*. Dublin: Trinity College, Centre for Language and Communication Studies.

Little, D., Ó Murchú, H., & Singleton, D. (1985a). *A functional notional syllabus for adult learners of Irish*. Dublin: Trinity College, Centre for Language and Communication Studies.

Little, D., Ó Murchú, H., & Singleton, D. (1985b). *Towards a communicative curriculum for Irish*. Dublin: Trinity College, Centre for Language and Communication Studies in association with Bord na Gaeilge.

Norton, S. (1984). Bridging the formal and informal fields of Irish language learning at the primary level. In D. M. Singleton & D. G. Little (Eds.), *Language learning in formal and informal contexts* (pp. 125-136). Dublin: Irish.

Ó Murchú, M. (1984). Irish and English now. *Ireland Today: Iris na Roinne Gnothai Eachtracha/Bulletin of the Department of Foreign Affairs, 1001*, 11-13.

Ó Riágain, P., & O Gliasáin, M. (1984). *The Irish language in the Republic of Ireland 1983: Preliminary report of a national survey*. Dublin: Institiuíd Teangeolaíochta Éireann.

Richterich, R. (1973). Definition of language needs and types of adults. In *Systems development in adult language learning by adults* (pp. 29-88). Strasbourg: Council of Europe.

ADAPTING ACTFL PROFICIENCY GUIDELINES TO RUSSIAN: PROBLEMS, APPLICATIONS AND IMPLICATIONS

Irene Thompson
George Washington University

1. Some problems in the adaptation process

The availability of government testers to train the initial contingent of academic testers in Russian made it possible for a group of trained individuals to begin work on the Russian guidelines in 1984. Unlike the other less commonly taught languages Russian, an Indo-European language, faced no special challenges in developing language-specific guidelines from a generic starter kit. This adaptation process could, perhaps, be best characterized by a conflict between the desire to make the level descriptions come to life through a variety of examples and the desire to preserve the global character of these descriptions.

The process of adaptation was not without some uneasiness caused by a conflict between the desire to make the Russian guidelines conform to those in French, German, and Spanish, and the need to include in them references to features that are unique to Russian. These features were related to content/context, accuracy, and the lack of provisions in the generic guidelines at the Novice Level for the learning of another alphabetical system.

With respect to content/context, the committee members felt that they had to correct the West-European and adult/professional bias in favor of contexts in which American students in the Soviet Union would be most likely to find themselves, and to provide examples of topics that Americans would be most likely to discuss with Russians particularly at the Advanced and Superior levels. It was obvious that many of the survival situations mentioned in the French, Spanish, and German guidelines could not be applied to Russian because they would either simply not occur in the USSR or would be characterized by an entirely different script.

Far more serious problems presented themselves in the area of accuracy. For example, the provisional guidelines in the Advanced-Plus description referred to lack of accuracy in the following way: "Areas of weakness range from simple constructions such as plurals, articles, prepositions, and negatives to more complex structures such as tense usage, passive constructions, word order, and relative clauses". Such specificity caused problems for Russian where additional grammatical categories such as pronominal, adjectival and nominal declensions, aspect, modality, verbs of motion, and prefixation, among others, present significant difficulties for the learners.

The removal of all references to specific structures in the revised guidelines of 1986 made the subsequent revision of the Russian-specific guidelines easier, for the committee members no longer felt constrained by the imposition of developmental hierachies for grammar more characteristic of less inflected languages. As a result the revised Russian guidelines in the description of the Advanced-Plus speaker refer to cases, aspect, mood, word order and the use of particles.

Although all members of the Russian guidelines committee had been trained in the administration of the oral proficiency interview and all were experienced teachers of

Russian, there was some uneasiness in positing a developmental hierarchy of acquisition of grammatical, discourse, sociolinguistic and pragmatic features based on observation and experience rather than on solid psycholinguistic research. It was felt then and still is that the availability of large amounts of data from taped oral interviews in many different languages should provide the impetus for psycholinguistic research into characteristics of speech at different levels of proficiency, partly along the lines suggested by Byrnes (1987) and Canale (1986). The results of this research may help to re-examine and re-evaluate some of the statements in the current versions of the proficiency guidelines with regard to various acquisition hierarchies.

Finally, the lack of accomodation in the provisional reading/writing guidelines for learning to recognize/produce Cyrillic letters caused the committee members some concern. While some transfer can be made from Western European languages when it comes to recognizing/producing Cyrillic letters, most of them represent a different pattern of letter-sound correspondence or have different shapes altogether. In addition, the printed and longhand versions of the letters look quite different. Thus, the learner has to go through a training period before s/he can recognize/produce Cyrillic script. The provisional guidelines, however, described the Novice-Low reader as having no functional reading ability, whereas the Novice-Mid reader was described as already being able to read highly contextualized words or cognates within predictable areas such as names, addresses, dates, street signs, building names, short informative signs, etc. It was felt that there was a discontinuity between these two subranges which did not reflect the process of learning to recognize Cyrillic.

The problem was solved by having the Novice-Low reader in Russian recognize some letters of the Cyrillic alphabet in printed form and a few international words and names. By contrast, the Novice-Mid reader could identify all letters of the Cyrillic alphabet in printed form and some contextualized words such as names, public signs, etc. Finally, the Novice-High reader could identify various typefaces in printed form or in longhand as well as highly contextualized words/phrases/sentences on maps, buildings, in schedules, documents, newspapers and simple personal notes. In this manner, the Novice level was felt to represent the gradual stages in learning to recognize the Cyrillic alphabet.

2. Applications

The recommendations of the National Committee on Russian Language Study (1983) which called for the development of a common metric and its use to set standards for Russian language study helped pave the way for the introduction of the proficiency guidelines and the oral proficiency interview into the Russian field. The response has been generally quite positive, and it is fair to say that there has been a lot of activity involving the guidelines and the oral interview test in the last two years. The following deserve mention:

(1) A dozen or so testing workshops later there are now thirteen certified oral proficiency testers and two academic tester trainers in Russian, which ends dependence on the U.S. Government for training.

(2) Curriculum workshops geared to teaching for proficiency are now being offered to secondary and postsecondary Russian language teachers throughout the academic year and especially during summers at various locations nationwide.

(3) Videofilms of oral interview tests at all levels were developed at Middlebury College under a grant from SSRC for use in tester training.

(4) ETS has developed an Advanced Russian Listening and Reading test based on ACTFL Listening and Reading Guidelines. The test reports raw scores and/or proficiency ratings from Intermediate High (ILR 1+) to Superior (ILR 3) or higher. The test was validated against DLI ratings and normed on about 500 university students with three years or more of Russian study.

(5) Major Russian overseas programs such as CIEE and ACTR use the ACTFL oral proficiency interview and the ETS Advanced Listening/Reading Test for pre- and post-program evaluation of participants. Data is being collected to update Carroll's (1967) study with respect to Russian.

(6) Some institutions have introduced graduation requirements for undergraduate and graduate majors in Russian in terms of proficiency levels in various skills. Other institutions are using the oral proficiency interview to screen prospective TA's. Many institutions are re-evaluating their language courses by setting course objectives in terms of proficiency levels in various skill combinations.

3. Implications

On the practical side there is little doubt that the proficiency guidelines have succeeded in injecting some vitality into the Russian language teaching field by offering both a framework for program planning and instruments for assessing student progress.

On the theoretical side, there is little empirical research on the acquisition of Russian as a foreign/second language to back up many of the statements in the Russian guidelines. This lack of research may be partially due to the relative recency of the Russian guidelines. As of now we lack:

(a) empirical studies to determine differences in specific aspects of linguistic performance across major boundaries, of specific linguistic features that distinguish planned from unplanned discourse, of features that characterize presence or absence of oral paragraphs, of differences in pragmatic and discourse strategies at various levels, and of hierarchies in the development of control over various structural features;

(b) studies of inter-rater reliability between government testers who use the ILR guidelines and academic testers who use the ACTFL Guidelines;

(c) investigations of differences between the way meanings are negotiated in real conversations and the way they are negotiated in the oral interview since there is a real possibility that certain conversational principles may not be observed in the oral interview;

(d) examinations of differences in testing one's own students as opposed to testing someone else's with regard to elicitation and rating;

(e) investigations of differences in elicitation and scoring by native as opposed to nonnative interviewers particularly at the Superior Level;

(f) examinations of the validity of the developmental sequence outlined in the receptive skills guidelines;

(g) studies of inter-rater reliability at different levels of proficiency;

(h) a better definition of functions which are presently a hogde-podge of several different things.

4. Conclusion

The development of language proficiency guidelines is an ambitious attempt to capture the most salient features of interlanguage at various points in its development and to describe them in a few well chosen sentences accompanied by a few carefully selected examples. For such an attempt to be realistic, it can only be viewed as a dynamic process of constant refinement based on our constantly expanding understanding of interlanguage development. In our attempts to disect the interlanguage continuum from total lack of competence to native-like performance we can take comfort in the fact that interlanguage development is characterized by a certain amount of invariance, or natural order, in the acquisition of various linguistic structures. Thus, if we knew what the acquisition curves for various linguistic structures for a particular language are, we could better describe what learners at certain levels can do and with what degree of accuracy. This is a formidable task which can never be complete. As a consequence, the Guidelines will always reflect a stage in our imperfect understanding of the dynamics of interlanguage.

Although originally designed as a practical tool for assessing language competence, the proficiency guidelines of today and tomorrow may ultimately have their greatest impact not so much on language teaching and testing practices as on psycholinguistic research generated by a desire to find answers to the many questions that the Guidelines have raised.

References

American Council on the Teaching of Foreign Languages. (1982). *ACTFL provisional proficiency guidelines*. Hastings-on-Hudson, NY: Author.

American Association for the Advancement of Slavic Studies. (1983). *Russian Language study in the United States: Recommendations of the National Committee on Russian Language Study*. Stanford, CA: Author.

American Council on the Teaching of Foreign Languages. (1986a). *ACTFL proficiency guidelines*. Hastings-on-Hudson, NY: Author.

American Council on the Teaching of Foreign Languages. (1986b). *ACTFL Russian proficiency guidelines*. Hastings-on-Hudson, NY: Author.

Byrnes, H. (1987). Second language acquisition: Insights from a proficiency orientation. In H. Byrnes & M. Canale (Eds.), *Defining and developing proficiency: Guidelines, implementations, and concepts* (pp.107-131). Lincolnwood, IL: National Textbook.

Canale, M. (1986). Language assessment: The method is the message. In D. Tannen and J.E. Alatis (Eds.), *Languages and linguistics: The interdependence of theory, data, and applications* (GURT 1985), 249-262. Washington, DC: Georgetown University Press.

Carroll, J.B. et al. (1967). *The foreign language attainments of language majors in the senior year: A survey conducted in U.S. colleges and universities*. Cambridge, MA: Graduate School of Education, Harvard University.

Educational Testing Service. (1986). *Russian proficiency test*. Princeton, NJ: Author.

THE APPLICATION OF THE ILR-ACTFL TEST AND GUIDELINES TO INDONESIAN

John U. Wolff
Cornell University

Indonesia is the fifth most populous nation in the world. With more than 165 million inhabitants it is next behind the United States in population and far ahead of France, Japan, Germany, and Italy, just to name a few of the countries whose languages are most widely studied by students in Indonesia. Indonesia is important to the United States economically, militarily, diplomatically and (I would insist) culturally, as well; yet the total enrollment in Indonesian is ludicrously small, amounting to less than 100 in any single academic year. Indonesian is the official language of a country, where neither English nor any other widely-known Western language is spoken by any significant portion of the population. There are good pedagogical materials available for Indonesian, the traditions for teaching Indonesian are positive, and the language does not offer horrendous difficulties to the learner. In short, there is every reason for the study of Indonesian to increase in this country, and there is every reason to encourage enrollment. For this reason, it is not amiss to discuss the proficiency-testing movement in relation to Indonesian instruction and testing.

By proficiency testing we mean an oral interview which was developed in the US government services and has since been applied to second-language testing outside of the government, especially in institutions of higher learning in this country, the purpose of which was to determine the testee's ability to handle oral communication.* These interviews are open ended conversations in which the examiner determines the proficiency of the person being tested by means of a series of questions or statements to which the testee must react or by means of requests to perform certain tasks in the language being tested. The testee is given a rating ranging from novice through intermediate, advanced, and superior, with pluses and minuses to clarify the testee's proficiency more precisely. In rather loose terms, novice refers to a student at the level of one who can do little with the language and whose repertoire is confined to recitation of individual vocabulary items; one who is at the intermediate level has some of the basic structural features but certainly not much control; one at the advanced level has much of the structure and good control (especially at the advanced plus level); whereas one at the superior level is a person who has worked for years on the language and is able to manipulate its expressive capabilities on a variety of subjects and express himself or herself with ease and accuracy with few structural errors. Since language learners show comparable characteristics in their handling of the second language at various levels of proficiency from language to language, it has been possible to develop generic guidelines which state the characteristics of learners at each level and which are valid for all languages. These generic guidelines are then made more precise by language specific guidelines, which outline in some detail what types of behavior characterize learners of the specific language at each level, discussed in terms of specific grammatical forms, constructions, or other features which are important for the language concerned. It is these specific guidelines which enable different examiners at different times and with different populations to give consistently accurate ratings.

---

*In recent years, tests of ability in listening, writing, and reading by similar methods have been developed for some of the major languages, but we will not be discussing this kind of testing in this paper.

These language specific guidelines have been developed for the major Western languages, and for the non-Western languages guidelines have been developed (or have been worked on for some years now) for Chinese and Japanese. For a rarely taught language such as Indonesian, it would be well if such guidelines were developed for three reasons. First, Indonesian is an important language for the reasons mentioned in the opening of this paper, and although enrollment is miniscule, a very high portion of the students of Indonesian make careers in fields in which a knowledge of Indonesian is an absolute requisite. Thus, it is crucial that the teachers of Indonesian be able to assess success of what is being done in the classroom. Second, it is necessary to ascertain a student's proficiency to determine qualification for admittance to study programs in Indonesia and at the career level, to determine an applicant's suitability for employment in a field in which a knowledge of Indonesian is vital. Finally, there is a reason which is applicable to all foreign-language teaching: testing of oral proficiency has had a salubrious effect on the language teaching profession in those institutions in which proficiency testing was adopted seriously because the expectations of student performance were based on a model of functioning in the speech community. Since what is tested is not mastery of a set of lexical items or grammatical patterns but the ability to perform, the teachers are led to adopt classroom procedures which consist of exercises in practicing communicative behaviors rather than exercises in inculcating an understanding of grammatical forms. Few who have undergone training in proficiency testing and who have mastered the guidelines have failed to gain rich insights into what language teaching should aim for and how one might achieve this aim.

In short, there is every reason to develop guidelines for proficiency testing in Indonesian. We are, however, still in the very early stages of developing such guidelines, and there is a great deal of spadework which remains to be done before valid guidelines for Indonesian can come into existence. One piece of spadework needs to be done with the generic guidelines to make them more generally applicable. The other spadework is in Indonesian itself. Let us talk briefly on what must be done with the generic guidelines. Although Indonesian has no unique features which resist measurement by a metric common to Western languages, it is true that Indonesian grammar revolves around very different principles, and generic guidelines must avoid presupposing any of the specific principles which characterize Western languages. However, generic guidelines which are expressed in terms of behavioral patterns will hold. That is, generic guidelines are applicable across the board if they are stated in terms of control of the discourse structure such that the interlocutor is able to understand sequencing and identities of the various subjects or objects of the actions, or if they are stated in terms of control of communicative strategies so that the interlocutor can follow the speech act being engaged in and the person tested can follow the speech acts of the examiner. In short, for all languages part of proficiency testing is ascertaining the ability of the student to clarify the parameters of the speech event in which he is engaging, and guidelines which refer to this goal of testing are applicable everywhere.

Let me now go back to spadework which must be done in developing guidelines specifically for the testing of Indonesian. For Indonesian problems of developing guidelines must be attacked on two fronts. First, is the determination of which features of grammar, lexicon, and organization at the sentence and larger level are indicative of speaker ability at each stage of proficiency and the second is the determination of the testee's ability to make use of the stylistic resources by which Indonesian enables the members of the community to function in accordance with the ethics which underly peculiar Indonesian cultural institutions. The student of Indonesian has the task of learning not only on the plane of grammar, lexicon, semantics, and functional sentence perspective but also on the plane of style, register, and sociolinguistic rule if he is to learn to function in the Indonesian speech community; and proficiency testing in Indonesian must address both planes.

Before we go ahead, it would be well to make a few remarks on the importance of sociolinguistic rules for Indonesian, where the expectations of the members of the speech community are at variance with behavior that is normally within the ken of the Western student, and where on the Indonesians' part there is not a great deal of understanding or tolerance for deviance from expected behavior. Whereas a linguistic faux pas may be the cause of amusement or discomfort in Europe, in Indonesia it can be the cause of serious tensions. Unfortunately, Indonesian is one of those languages in which almost every time one opens one's mouth, a strong statement is made about human relations, statuses, or what kind of a person everyone involved in the conversation is. Indonesian is heavily dominated by Javanese culture, which revolves around human relations, and the Javanese ethnology of communication or method of interaction has become the national norm. In short Javanese, and thus Indonesian by virtue of being dominated by Javanese, has clear notions of how society should be ordered and accordingly has created an ethic which is enforced by strict adherence to sociolinguistic rules. Although the repertoire of Javanese does not spill over into Indonesian in its total complexity and richness, there is enough complexity there to make unequivocal statements about human relations and put into effect these rules for interaction. These features involve not only terms of address and reference, but also they involve demands for a certain amount of indirection, circumlocution and reference by inference. Failure to adhere to these rules of indirection and circumlocution also makes a statement of social identity. Thus these matters are part and parcel of elementary Indonesian language instruction, as much as morphology, syntax, lexicon, and the semantic system of sentence perspective. These matters thus all must have a place in testing, and a set of guidelines which enable examiners to test students in a way that is consistent and accurate must make specific statements of how control of these matters manifest themselves at various stages or proficiency levels.

In developing proficiency guidelines for Indonesian, the first step which the community of specialists in the teaching of Indonesian must take is to gain skills in testing of a Western language for which there is a tradition of testing and for which there are fairly effective guidelines in existence. On the basis of this training and skills derived therefrom, Indonesian teachers can prepare themselves to interview students for whose proficency level one has a fairly good intuitive level. These interviews should be conducted with students at all levels and the interviews should be transcribed. The purpose of these trial interviews is two-fold: (1) develop a set of testing routines which is applicable to the peculiar Indonesian cultural setting and can elicit the routines which demonstrate a mastery of sociolinguistic routines; (2) provide data of the types of features which mark the speech of students at each level. On the basis of these interviews (and there will necessarily be many dozens) a set of preliminary guidelines can be worked up. These then will have to be distributed throughout the community of specialists in Indonesian pedagogy who will then apply them further and work on refining them.

It seems almost quixotic to propose such elaborate and complex processes for proficiency guidelines in a language so rarely studied, and for which there is no large cadre of teachers who could be taught to administer a proficiency interview. However, I would submit that even though proficiency testing for students in Indonesian around the country is not a practicable goal, guidelines specific to Indonesian are nevertheless needed in order to prepare a test which can be practically administered throughout the country. Any such test would probably have to be pre-recorded with questions worked out in advance. It is possible to prepare a pre-recorded test which approximates or replicates a proficiency interview only if it is based on experience with a proficiency interview. In short, the proficiency guidelines must be developed as a step in the process of developing a pre-recorded "proficiency mode" test which aims to replicate or approximate the oral interview. Thus, even though we do not foresee extensive testing with proficiency interviews for Indonesian, the guidelines and training which will enable specialists in teaching Indonesian in this country to conduct proficiency interviews must be developed as a step on the way to developing a pre-recorded "proficiency mode" test which can be administered to the widely scattered population of students in Indonesian in various institutions in this country.

Summary of the panel discussion on THE EXTENSION OF LANGUAGE PROFICIENCY GUIDELINES AND ORAL PROFICIENCY TESTING TO THE LESS COMMONLY TAUGHT LANGUAGES

Robert Vosika and Timothy Buckwalter
Indiana University

The discussion among the panelists centered on their experiences in applying the ACTFL Guidelines to their respective language areas, with special emphasis given to the following topics:

Specific versus generic guidelines – Panelists discussed the tension between the selection of representative versus descriptive items in the writing of the language-specific guidelines. Irene Thompson suggested that a delicate tension exists between too much and not enough specification in the grammatical descriptors of level guides: while there is the need to provide a sufficient number of examples to make each level "come to life", there is also the danger that too much specification may invite use of the guidelines as a list of discrete items to be tested. Ernest McCarus proposed using the degree of accuracy in structure rather than discrete linguistic features, vocabulary or structures in determining the level of proficiency. John Wolff suggested that the generic guidelines reflect discourse structure and that the language-specific guidelines reflect grammatical structure. Susan Gass asked whether it would be appropriate to develop future language-specific guidelines using specific mention of descriptors, recognizing the bias created by their use in the generic guidelines. Irene Thompson responded that experience has shown that OPI trainees use these descriptors as examples rather than as a list or inventory. It was concluded that there has not been enough psycholinguistic research on which to base these descriptors.

Needs versus tasks – A second issue that was raised concerned the question of whether needs would be a better basis for defining proficiency levels rather than using a description of the specific language. For Jos Nivette proficiency guidelines must be based on the students' needs rather than on time or effort required to attain proficiency, the nature of the language itself, or the means or methods used to get there. David Singleton elaborated on this point in the context of Irish, where it was felt that learners' interests and expectations had to be recognized in terms of the kinds of tasks that they would be asked to perform in proficiency tests, rather than using teacher-perceived survival needs as the point of departure. He noted, furthermore, that any assessment of survival needs for a language whose speakers also have a native-like competence in a second and widely spoken language, will run into problems of artificiality. Albert Valdman, using Haitian Creole as an example, reinforced the point of looking at language needs in a diglossic context. A note of caution was raised by Lyle Bachman, who remarked that by dwelling on the specificity of languages we run the risk of carrying the argument to its extreme conclusion: that there is a specific language for each learner, and that this is obviously not where the profession wants to go. A problem that must be dealt with, this participant added, is the inherent dilemma that is involved in defining proficiency in terms of actual language tasks.

The issue of diglossia – In a discussion of the inherent difficulties involved in assessing proficiency in languages that exist in a diglossic context, Ernest McCarus pointed out that testing for sociolinguistic appropriateness in Arabic presents problems for academia, given the fact that most university programs focus solely on the literary language -- Modern Standard Arabic -- whereas Colloquial Arabic is the proper language medium for meeting basic survival needs. Rather than choosing to ignore the colloquial dialects, or to write two separate guidelines -- one for Standard Arabic and one for a major dialect, such as Egyptian Colloquial -- McCarus proposed that Arabic be treated as a living language, with one set of guidelines designed to measure appropriateness of language usage: use of Colloquial and Standard Arabic according to how the educated native speaker uses both. On the other hand, Albert Valdman noted that for Haitian Creole the ideal criterion of measure is not the Educated Native Speaker, but rather, the non-ENS, since the primary speakers of Creole are illiterate peasants. Assessment at the lower levels of proficiency must take into account the native speaker's use of pseudo-French expressions, whereas, attainment of advanced proficiency implies bilingual competency and the ability to pepper one's Creole with French expressions and structures. Valdman added that this situation also held true for all African languages which exist in a diglossic relationship with French.

The sociolinguistic/cultural issue – A major issue that was discussed concerned the assessment of the cultural and sociolinguistic component of foreign language proficiency. Ronald Walton pointed out that the typical OPI may not adequately address sociolinguistic factors of languages whose cultures are vastly different from those of European language communities. He added that this problem may have been addressed by the cultural guidelines of the ACTFL Provisional Guidelines, which were later removed for lack of an adequate theoretical framework. He also remarked that appropriate cultural linguistic behavior, for a foreign student of Chinese, is not necessarily that of the host native speaker, who may expect appropriate foreign behavior rather than an attempt to imitate behaviors of the hosts. Albert Valdman noted that proficiency in Haitian Creole must include an assessment of the total cultural matrix in which the language is spoken. He pointed out that to attain the competence of an Intermediate speaker, one would have to know a minimum of 500 proverbs, and that knowledge of these proverbs implied a knowledge of the role of vaudoun in Haitian culture. Irene Thompson remarked that adjustments in the content of the Russian guidelines had to be made, and that the entire set of situation cards had to be rewritten due to their Western European bias. Finally, John Wolff emphasized that the guidelines for languages whose culture differ considerably from Western cultures will, per force, look quite different from European language guidelines.

Multi-component versus global ratings – In connection with the discussion on the cultural and sociolinguistic components, several panelists expressed concern that a global rating alone was not considered adequate for student evaluation, and that there was a perceived need for providing diagnostic information. In this regard, Sally Magnan offered a multi-component model currently being developed for African languages, which gives a complex profile of 36 bits of information, or 36 separate evaluations, each one based on a 3-point scale of degree of mastery. This model distinguishes four language modalities: aural/oral and reading/writing. Furthermore, it includes the assessment of three aspects of language not measured by the ACTFL OPI: social/cultural, lexical/traditional, and emotive/pragmatic. In addition, this model required an interview format expanded beyond the ACTFL model that would include many more interactive exchanges in "pseudo-authentic" cultural contexts. However, she granted that there exist problems inherent in any model which is very comprehensive in theory and very detailed in its evaluations. Ernest McCarus also favored using a multi-component score in addition to a global one for Arabic in order to provide diagnostic information, such as linguistic accuracy, communicative proficiency, and appropriateness of style. Irene Thompson noted that the OPI is in fact used in Russian as a diagnostic tool for internal purposes.

Future research – In the concluding part of the session a number of suggestions were put forward for future research. Lyle Bachman suggested that the profession should not proceed with the assumption that it will always have *x* number of scales across the board for all skills, for all languages, and that empirical research will be needed to determine the optimum number of scale levels. Irene Thompson called for psycholinguistic research on which to base the level descriptions themselves, and recommended use of the guidelines as a heuristic tool to that end. John Wolff suggested that direct testing may not be the most efficient means for assessing oral proficiency in languages with extremely low enrollments, and that alternative methods, such as semi-direct tests, may prove to be more practical in dealing with these languages. And, finally, Lyle Bachman emphasized the need to examine whether the ACTFL Guidelines are, in fact a common metric for assessing proficiency in different languages, and the need to look further into the assumption that some language features are amenable to measurement by a common metric, whereas others are not.

TESTING LISTENING COMPREHENSION IN THE CONTEXT OF THE ACTFL PROFICIENCY GUIDELINES

Dan Douglas
Iowa State University

This paper is a discussion of issues related to the testing of listening comprehension in the context of the 1986 ACTFL Proficiency Guidelines. After a review of current research on listening comprehension and development work on listening tests, four issues are singled out for more detailed consideration: the meaning of "context" in listening testing, the concept of criterion-referenced tests, the notion of specific purposes in testing, and the use of technology in listening tests. A number of research issues are discussed.

1. Introduction

Listening comprehension has been taking on growing importance in foreign language learning research. Asher's demonstration (Asher, 1969) of Total Physical Response techniques has helped focus thinking on the centrality of listening comprehension. Judith and Norman Gary point out that "in recent years there has been an accumulation of research - both of a theoretical and of an experimental nature -which strongly suggests that there should be much more emphasis on the comprehension aspects of language in foreign language teaching than is usually the case." (Gary & Gary, 1981, p. 1)

"However," they continue, "an examination of recently published textbooks makes clear that this research is having little impact in the average classroom." The same may be said of listening comprehension testing. Although a great deal of research has been conducted in recent years on the listening skill, and its central place in initial second language learning and its influence upon the development of the other language skills has become better understood, listening comprehension testing has been, and remains, somewhat the "quiet corner" in the testing field.

A number of researchers have provided insights into the complexity of the active listening skill: Richards (1983) considers "message" and "medium" factors to arrive at a list of 33 "micro-skills" required for conversational listening, and a list of 18 micro-skills relevant to academic listening; Chaudron and Richards (1986) show that "macro-markers" (those that highlight information in a lecture and its sequencing or importance) are more important to comprehension than are "micro-markers" (which indicate links between sentences in the lecture); Glisan (1985) provides evidence for the importance of first- and second-language word order in the comprehension of key sentences in texts; Dunkel (1986) deals with the strategic processes of listening comprehension; and Brown (1986) reports on experimentation on the influence of context in listening comprehension.

Powers (1986), in a report on a survey of the academic demands related to listening skills, outlines variables related to stimulus material, speaker and context as they affect listening comprehension, and identifies nine specific skills as very important: identifying major themes or ideas, identifying relationships among major ideas, identifying the topic of a lecture, retaining information through notetaking, retrieving information from notes, inferring relationships between information, comprehending key vocabulary, following the spoken mode of lectures, and identifying supporting ideas and examples.

In short, one finds in current research a view that listening comprehension is a complex, active skill that is basic to and influential in foreign language learning. What evidence is there that this lively research interest has carried over into testing? What's going on currently in listening comprehension testing?

2. Current work in listening comprehension testing

In order to provide a brief overview of current work in listening test development and research, I employ the following fuzzy categories: "Listen and Give the Right Answer" tasks, "Reduced Redundancy" tasks, "Interaction" tasks, "Repetition/Imitation" tasks, and "Media-transfer" tasks. No rigorous theoretical implications are made by these categories; in fact, I would like to avoid, in this paper, getting mired in sorting out the many orientations, approaches and bases which we find in the literature currently: e.g. notional/functional, the proficiency movement, task based, content based, performance based, pragmatic and communicative. As Alan Davies pointed out some years ago, our field of applied linguistics is much better at analysis than at synthesis.

2.1 Listen and get the right answer tasks – Here I include work being done in the area of standardized listening tests--the TOEFL-type listening comprehension format, where, although the length and complexity of the stimulus varies, the main task is to listen and then answer questions about the text. The research and development work in this category have been focused recently on item response theory (e.g. Woods & Baker, 1985) and the problems of developing equivalent forms and of adapting items to individual candidates' proficiency levels. As reported by Stansfield (1986), work has also been progressing in the TOEFL program on investigaing the effects of manipulating the stimulus material by including such features as pauses, expressions of hesitation and uncertainty, false starts and appropriate background noises (e.g. talk and clanking dishes in a restaurant situation).

2.2 Reduced redundancy tasks – In these tests, the primary focus is to give candidates an opportunity to show that they can process language in spite of reduced redundancy, effected usually by mutilating the text in some way, or by adding "noise" to the transmission. Candidates may demonstrate proficiency either by "giving the right answer" in spite of the noise or mutiliation, or by filling in the missing or masked portions of the text. The most well-known examples of reduced redundancy tests include the "noise test" (Gradman & Spolsky, 1975), produced by adding white noise to the transmission of stimulus sentences which candidates have to reproduce in writing; and partial dictation and listening cloze tests (Henning, Gary & Gary, 1983), where candidates are given a mutilated written text and have to fill in the gaps by listening to an unmutilated spoken version.

2.3 Repetition/imitation tasks – Here I would include work on dictation tests (Fouly & Cziko, 1985), particularly the recent work of Cziko and his colleagues in studying the scalability of dictation test results, based on chunks of increasing length; the Integrative Grammar Test (Bowen, 1976), an interesting example of the use of "natural" reduced redundancy (reduction, assimilation and contraction) to test grammar; and elicited repetition (Perkins, Brutton & Angelis, 1986), where students are asked to repeat sentences as a demonstration of their control over specific syntactic structures. Mydlarski and Paramskas (1985) have developed a computerized template program to allow teachers to easily develop dictation exercises and tests in which students respond via the keyboard.

2.4 Interaction tasks – In tests involving interaction tasks, candidates demonstrate comprehension by providing an appropriate response to a conversational stimulus. Formats of interaction tests range from oral interviews, which frequently make provision for a comprehension score, to multiple-choice formats where candidates are asked to choose an appropriate response. One interesting interaction test is that reported by Friel (1984), called the Discourse Cooperation Test, which employs Grice's maxims of conversational cooperation. In the test, a situation is established, such as an automobile accident, and the examiner initiates the interaction by "flouting" one or another of the Gricean maxims. The candidate demonstrates comprehension of the flouted maxim by choosing an appropriate response from five options, each of which is related to one of the maxims, plus a "neutral" response. Computerized interactive audio/video programs, such as the "Montevidisco" project at Brigham Young University (Gale, 1983; Schneider & Bennion, 1983) also offer an interactive listening comprehension testing capability. For example, the student is confronted, by means of videodisc, by a native of a Mexican village and asked whether she would like a tour guide. She demonstrates comprehension by choosing an appropriate response, which may include a request for repetition (which, owing to the "magic" of the videodisc technology, is readily supplied). Interactive audio/video applications will be discussed in more detail in a later section.

2.5 Media-transfer tasks – Here I would place a variety of listening comprehension test activities which involve the candidates' abilities to transfer between two or more media -- aural, written, pictorial, diagrammatic--to demonstrate comprehension. For example, the Michigan English Language Battery Listening Comprehension subtest presents students with a map of a picnic area while providing information aurally for the interpretation of the map. Comprehension is shown by the students' ability to answer questions about the map. The British Royal Society of Arts examination requires students to complete a short schedule by listening to an announcement; the British Joint Matriculation Board examination has a section where students hear an explanation of the technical labels on refinery gas lines and then have to fill in blank labels based on further aural information. Also in Britain, the new Test in English for Educational Purposes (Emmett, 1985) involves candidates in a preliminary reading task, followed by a related lecture on which comprehension questions are asked. Candidates then write down some dictated notes and complete a writing task based on what they have read and heard. As a final example of this integrated transfer type of test, Madden and Samuda (1985) report on a task- based test in which candidates listen to a conversation between two people planning their day's activities. The students have a map on which they are to mark the route of the planned excursion, and a small table to fill in with information about times and places. This is followed by reading, writing and speaking tasks, all integrated around not only the topic but previous input.

This brief review suggests to me that current work in listening comprehension test development is moving in the direction, first, of tests which are more efficient - adaptable to individual abilities; second, tests which are both integrative (in the technical sense of testing more than one skill or ability at once) and integrated (in the sense of listening comprehension tasks feeding and being fed by other skills); third, tests which challenge learners to employ varying strategies to deal with varying conditions of transmission noise, text mutilation or differing media; fourth, tests which are contextualized; and finally, tests which require students to apply understanding gained through listening to a variety of tasks involving other components of language proficiency. In light of this understanding of current work, I would like to turn now to a consideration of the ACTFL Proficency Guidelines and suggested directions for test development in the ACTFL context, particularly those which came out of a symposium and working group on testing the receptive skills reported in the September, 1984, issue of *Foreign Language Annals*.

3. The ACTFL Guidelines

Canale (1984) has suggested two overriding principles in considering the testing of reading and listening proficiency: 1) attempt to elicit the best performance from test takers by presenting tasks that are fair, important in themselves and interesting in themselves; and 2) attempt to provide test tasks that can reveal to test takers and educators clear, rich relevant and generalizable information. More detail was provided in a synthesis produced by Canale, Child, Jones, Liskin-Gasparro, and Lowe (1984) in which the following recommendations were made:

1. A focus on criterion-referenced proficiency testing that would provide both profile scores and fine-grained diagnostic information for a variety of purposes.

2. Test organization based on thematic unity (as opposed to coverage simply of structural points or skill areas); and the four ILR oral interview stages (warm-up, level-check, probe, and wind-down).

3. Item specifications should reflect the interrelationship among performance level, text type, reading/listening function, and reading/listening strategy (See appendix A for a table outlining this interrelationship).

4. An adaptive administration procedure should be devised to provide an efficient, accurate and affectively rewarding test administration.

5. Scoring criteria should be accuracy of comprehension, independence in dealing with text and speed of completion.

6. Item difficulty should be calibrated by means of reference to proficiency levels, pilot screening and on-going studies. Construct, predictive, concurrent and content validity studies should be carried out.

7. Learners, teachers and administrators should each receive useful feedback from the tests, such as raters' commentary and verbal summaries of group and individual performance.

These recommendations are reviewed by Dandonoli (1987), who then goes on to describe current reading and listening test development work within ACTFL in light of them. She provides an interesting and thoughtful discussion of the problems likely to be encountered in attempting to carry out the recommendations of the 1984 working group and her views will be referred to frequently in the remainder of this paper.

The current version of the ACTFL Proficiency Guidelines (1986, p. 15)--no longer "provisional"--"identify stages of proficiency...to allow assessment of what an individual can do and cannot do with language, regardless of where, when or how the language has been learned or acquired." Five levels are identified: Novice, Intermediate, Advanced, Superior and Distinguished. The first two stages have low, mid and high substages, while the third stage has an "advanced-plus" category. The generic description of the listening comprehension guidelines assumes "that all listening tasks take place in an authentic environment at a normal rate of speech using standard or near- standard forms." The Novice Level is characterized by topics referring to basic personal information or the immediate physical surroundings; Intermediate by topics on personal and social convention; Advanced, by topics beyond the immediacy of the situation dealing with factual information; the Superior Level by abstract topics in professional settings; and the Distinguished Level by comprehension of material on any topic, for a variety of needs tailored to different audiences.

Though it is beyond the scope of this position paper to critique the current form of the guidelines, I would mention some surprise that there is not closer adherence to the text and task specification outline provided by Canale and his colleagues (1984, reproduced here as Appendix A); I find it hard to discern the supposed developmental relationship among the five stages of the Guidelines. Apart from that criticism, which may be one merely of style and rhetoric, we must continue to remind ourselves that if the Proficiency Guidelines are to be truly said to identify stages of proficiency, empirical studies must be conducted to show that, for example, learners who can "understand main ideas and some facts from interactive exchanges" can also "recognize learned material and isolated words and phrases when strongly supported by context," but cannot do these things "beyond the immediacy of the situation," and so on up the scale. In other words, it is crucial that we obtain satisfactory information on the validity of the underlying construct of the Guidelines (cf. Savignon, 1985) before and in concert with further work on a national metric for foreign language proficiency.

Before turning to a detailed discussion of specific areas that seem to me to be both particularly interesting and nettlesome, I would like to just draw attention to the very lively debate on the proficiency guidelines and their impact that has been reported in various journals in the last couple of years (e.g. Savignon, 1985; Lantolf & Frawley, 1985; Bachman & Savignon, 1986; Kramasch, 1986; Lowe, 1986; and Schulz, 1986). The discussion has been undertaken with a view to refining the concept of proficiency from a communicative, interactionist point of view. The current Guidelines already reflect some of the points made in the debate and no doubt other points will emerge from the present Symposium. Interestingly, the result of this "refining" process has been not a narrowing of the notion of proficiency by winnowing out unnecessary components, but rather a widening of it to accommodate our growing understanding of how complex language in use is and how varied are the purposes for which we employ it.

4. Issues for discussion

I would now like to take up in turn four issues that seem to me to be of particular importance in our consideration of listening comprehension testing in the context of current research, test development and the ACTFL Proficiency Guidelines: criterion referenced testing, context in listening comprehension, language for specific purposes, and uses of technology.

4.1 Criterion referenced testing – We have seen (Canale et al., 1984; Dandonoli, 1987) that thinking toward the testing of receptive skills in the ACTFL context involves the concept of criterion-referenced testing. I am frankly uneasy about the meaning of "criterion-reference" in a language testing enterprise. The perceived advantage of criterion-reference is succinctly stated by Delamere (1985, p. 45): "...criterion-referenced tests make it possible for us to draw different kinds of inferences about a student's ability to use the language than norm-referenced tests do...For example, a criterion-referenced test might enable us to say that an individual learner spelled eight out of ten words correctly from a basic sixth grade vocabulary list and that his performance is therefore satisfactory..." Such a view highlights what I see as a flaw in the criterion-referenced position with regard to language testing, namely that it is beyond our present understanding of how language is structured and used to be able to say with any confidence that because a learner "spelled eight out of ten words correctly...his performance is therefore satisfactory." Or alternatively, that because a learner spelled only seven out of ten words correctly, his performance is unsatisfactory. As Denham (1975) points out, the establishment of a criterion level is often quite arbitrary; it may be that Cziko's term "edumetric" (Cziko, 1983) offers a solution to this problem--he would link linguistic criteria to eductional goals. However, it seems to me that the real difficulty is one reflecting the nature of languge use itself: there are so many ways, structural and strategic, to "get something done" with language, that it is currently beyond our ability to establish a domain of tasks which would add up to a criterion. Only a very narrowly focussed language for specific purposes context (cf. Widdowson, 1983) might allow for a reasonable criterion referencing. Cartier (1968) illustrates how this might be possible where "we know exactly where the student will go, what he will be studying there and what kind of work he will be doing afterward." In the much more general purpose, educationally oriented context the ACTFL Guidelines are directed at, however, and at our present state of linguistic understanding, an attempt at criterion-referenced language testing will necessarily be reductionist and, as Lantolf and Frawley (1985, p. 343) argue,"such tests construe A world; they do not reflect THE world." As Tollefson (1986) and Auerbach (1986) warn, a number of limiting biases are likely in materials and tests developed with explicit competencies in mind, selected, as they are, in the lack of any accepted methodology for discovering essential competencies. Discussion needs to focus on the meaning of criterion referencing in language testing and what kind of world our criteria would impose on learners.

4.2 Context and listening tests – Canale et al. (1984) have recommended that tests be organized around a unifying theme as a means of providing a more satisfying and motivating experience for candidates. Dandonoli (1987) supports this and points out that thematic unity also has the advantage of providing an authentic and meaningful context. Discussion needs to focus on what we mean exactly by "thematic unity" and by "context" and also on how context influences language performance. Brown (1986), for example, distinguishes between "external context," which constrains expectations for what is appropriate, and "discourse-internal context," which establishes shared information and shared conventions. She claims that discourse-internal context is harder for less able students to recognize and use. The use of internal context involves an "...ability to extrapolate from previous moves in the task and so to predict the expected range of the next move"(p. 300). Brown reports research on first-language listening comprehension in which subjects shared information on maps, and suggests that good and poor listeners differed not so much in their abilities to recognize inadequate messages, but more in "the ability to give appropriate feedback to the speaker when the message is inadequate" (p. 299). A refined notion is needed of what in the context is relevant for communication.

Faerch and Kasper (1986) list some "fairly uncontroversial" hypotheses in current models of first language listening comprehension, including the following:

1) Comprehension processes rely on three types of information: linguistic input, recipient's linguistic and socio-cultural knowledge, and contextual information.

2) The process of matching input with the recipient's knowledge can start either with the input or with the knowledge. In both cases, contextual information is used to support comprehension.

3) Different comprehension tasks focus on one or the other of the two processing directions.

4) It is the rule rather than the exception that the match between input and knowledge is less than perfect, resulting in "gaps," meaning that comprehension is typically partial rather than total.

Faerch and Kasper suggest that second language comprehension differs from first language comprehension only in that second language learners have more restricted knowledge of the second language, influenced by transfer from the first language, and have unrealistic expectations that comprehension must be total.

Overall, research such as that by Brown and Faerch and Kasper suggest that the influence of context on second language comprehension, while acknowledged to be important, is little understood as yet. Indeed, the work of such researchers as Tarone (1983), suggests that different tasks and/or contexts result in different second language performances by learners. Much more information is needed in this area of proficiency testing.

4.3 Language for specific purposes – "Proficiency" being defined as "the ability to use language for some real-life purpose (Dandonoli, 1987), one finds surprisingly little in the Guidelines about the specific purposes of language use around which the testing enterprise can be built. Discussion should focus, I think, on the issue of purpose, and how related issues of criterion-referencing, thematic unity and context might be dealt with. The table illustrating text and task specifications for reading and listening, provided by Canale et al. (1984), for example, related to work by Child (1987), is devoid of any reference to specific domains of language use. Somehow, it seems to me, provision must be made for incorporating the concept of specific purposes into the "general purpose"

ACTFL context. The work of Widdowson (1983) may be a starting place for discussion of this area.

Widdowson (1983) suggests that a difference between specific-purpose and general-purpose language teaching enterprises lies in the distinction between "purpose" as a descriptive term in specific-purpose operations, and as a theoretical term in general-purpose ones. What he means is that in general-purpose teaching, where the eventual use of the language being learned is not clearly discernable (the case being addressed by the ACTFL Guidelines, it seems to me), "purposes are specified by objectives, pedagogical constructs which seek to provide for the achievement of communicative aims when occasion arises after completion of the course" (p. 12). He would have us focus, in our specification of purposes of language use, on procedures for "making sense"--procedures being defined as "negotiating activities which interpret the directions provided and enable us to alter our expectations in the light of new evidence as the discourse proceeds" (pp. 40-41). How do we arrive at an understanding of these procedures so as to be able to include them in our teaching and testing? Widdowson suggests a problem-oriented approach to activating procedures, providing learners with the same type of procedural work as would be required in their various activities outside the language classroom (p. 91). Thus, it seems to me that procedural activity might legitimately be a focus of listening comprehension test devlopment, testing subjects' abilities to "make sense" of material related to specified domains of language use.

The problem that needs discussion, I believe, is the specification of domains of language use in the ACTFL context. What purposes (in the Widdowson "procedural" sense) do we have in mind when we speak of testing "proficiency?" Proficiency for what? It will not be enough to speak of "thematic unity," of "criterion-referencing" or of "text/task specification" without relating these notions to procedural activity associated with specific domains of language use.

4.4 Technology – There are two issues related to the use of technology in listening comprehension testing that might fruitfully be discussed: that of computer adaptive testing (CAT) employing item response theory, and that of interactive audio/video (IAV) techniques.

Computer adaptive testing – Dandonoli (1987) has suggested that computer adaptive testing procedures merit consideration in test development within the context of the ACTFL Guidelines, owing mainly, in her view, to the promise of greater efficiency. She cautions that there are some theoretical difficulties to be overcome, most notably that of the unidimensionality requirement of item response theory and that of a potential narrowing of the content range of a test owing to a reduction in the number of items allowed by the application of item response theory. Larson and Madsen (1985) argue that computer adaptive tests provide a number of advantages to the teacher, tester and student, among them its compatibility with computer assisted instruction, immediate feedback of results, ease of revision of test content, individualizing test items and flexibility in test length. They are particularly interested in the possibilities in item response theory, first, of presenting items within a range of difficulty close to the ability level of the examinee as a method of increasing test effectiveness, and second, of placing examinees on an absolute scale of proficiency, independent of the persons being measured and of the measure being used.

Canale (1986), in writing of the "promise and threat" of computerized adaptive testing attempts a balanced view of the advantages and problems, and in particular cautions that we may be in danger of trivializing our theories of comprehension and compromising their relevance to educational settings (p. 33). This echoes Tung (1986) who is also cautious that the requirement of unidimensionality in item response theory "raises serious doubts whether complex language behavior characteristic of real-life communication can be assessed with CAT" (p. 25). Canale also raises the issue of method effect in the use of CAT and calls for research into the potential problems associated with the mechanization of comprehension testing. Useful studies might be carried out comparing tests constructed outside the limits of item response theory with those adhering to its requirements. What desirable features of content are excluded or limited by item response theory, and how might these be incorporated in adaptive tests constructed by other means?

Interactive audio/video – The potential offered through the interfacing of the computer, videodisc and audio recorder, referred to as interactive audio/video, is almost overpowering. The "Montevidisco" project (Gale, 1983; Schneider & Bennion, 1983) has already been mentioned. It takes the student on a simulated visit to a northern Mexican town and in the process exposes her to real-life situations with natives speaking in Spanish. There are slightly different versions for male and female students owing to the "gender intensive" nature of Spanish. The overall effect is rather like that of a computer adventure game in which the player controls movement through a series of scenes and situations and must produce appropriate responses to them. In "Montevidisco" there are 28 major sequences (see the "map," Appendix B), each with several scenes, each scene having at least four options for the student to select. In all there are over 1100 possible "paths," and the program can be run a number of times, with the student encountering novel situations each time. Rowe (1985) sees IAV in the style of Montevidisco as the wave of the future, and calls for a NASA-like cooperative development effort to "deliver the goods" in an IAV mode. Other programs employing IAV include "Klavier im haus" (Schrupp, Bush & Mueller, 1983), a short program in German, following the problems of a couple moving into a new apartment; "Videodeutch" (Dahl & Luckau, 1985), another program in German which emphasizes non-verbal cultural aspects of language learning; and a pilot videodisc project in French (Eastmond & Mosenthal, 1985), which builds on the Montevidisco project by allowing student access to a main menu at any time to ask for help, change sound tracks from the original dialogue to one with a clear narrative, call up subtitles in either his first language or French, or call up a complete written text with translation.

What IAV offers the tester of listening is a capacity for the student to demonstrate comprehension by responding appropriately to natural, contextualized stimuli in the program by entering a multiple choice response on the keyboard, or by recording it orally on the cassette. The computer can easily make a diagnostic record of the responses, including requests for repetition, and the speed with which they were made.

One should also mention the somewhat more "one-way listening" possiblities offered by such technical innovations as connecting audio-cassettes to computers (e.g. PLATO IV at the University of Illinois) and digitally recorded voice data on floppy disks. Both types of audio material can be part of the computer program and can be randomly accessed. Listening cloze, notetaking, text reconstruction and deductive problem-solving tests can all be devised using the interaction of the computer and an audio device.

What needs discussion, of course, are the potential disadvantages of pursuing an

IAV approach to language teaching/testing. They include those already mentioned in connection with computer adaptive testing, since presumably one might want to employ CAT techniques in the construction of the IAV program. The issue of method effect is particularly relevant here, for what might run through a student's mind when confronted by the specter of videodisc player, audiocassette recorder, computer, keyboard, monitor, TV screen, microphone, and headphones? Natural communication? But there are also problems of the enormous cost and relative unavailability of IAV systems: the commercial viability of the videodisc has yet to be proven (indeed, one original manufacturer of the technology has already discontinued production). One project, in the face of this sort of problem, has moved to the use of the videocassette in interactive programs: "Autotutor" (Little & Davis, 1985) has the advantage of relatively low cost and the availability of huge amounts of video material adaptable to language teaching and testing. What it doesn't allow, of course, is immediate access to any part of the program, which is an essential component of "interaction." Research reported by Schrupp, Bush and Meuller (1983) suggests that without the interactive element, video loses some of its power as a teaching medium. It may be that we await further technological advances which will provide for the interactive capability of the videodisc without its expense and relative unavailability.

5. Conclusion

In this concluding section, I want to summarize suggestions for development work and validation studies in listening comprehension testing. Work in listening comprehension testing suggested in this position paper would be toward tests which reflect our understanding of the listening skill as complex and active. We seem to be moving in the direction of, first, tests which are more efficient in the sense of being more adaptable to individual abilities; second, tests which are both integrative in the technical sense of testing more than one skill at a time, and integrated in the sense of the listening skill being fed by and feeding other skills; third, tests which challenge learners to employ strategies to deal with varying conditions of noise, text mutilation, or media; fourth, tests which are contextualized; and fifth, tests which require learners to apply understanding gained through listening to a variety of tasks. Proficiency in listening comprehension would be related to empirical stages of second language development by reference to scales reflecting such development. Finally, test formats would be derived from technology associated with computer adaptive testing and interactive audio/video techniques. Validation studies will focus on relating test results with developmental stages, and on analyzing test content in relation to our understanding of the construct of languge proficiency.

In considering these tendencies in applied linguistics and language testing within the context of the ACTFL Proficiency Guidelines and the movement toward a national language proficency metric, I outlined four issues that seem to me to be both extremely interesting and thorny, and which might profitably be discussed today: criterion-referenced testing, context and listening tests, language for specific purposes, and uses of technology in testing.

1. Concerning the first issue, criterion referencing, we need, I think, to focus on problems arising from the nature of language itself--form and function relationships are just not as clear as we might once have hoped. It may be that we would want to consider

the advantages of interpreting language proficiency in terms of domains of behavior rather than in terms of criterion levels (Denham, 1975), but we need to consider carefully the effect of the resulting reductionism. To repeat the point made by Lantolf and Frawley (1985): criterion referenced tests "construe A world; they do not reflect THE world." Clark and Clifford (1987) have called for validation studies relating the test results to highly detailed, task-specific external criteria. It seems to me that it is precisely those highly detailed task-specific criteria that ought to form the basis of the tests themselves. The results would be a sampling of content domains and a more truly criterion-referenced set of measures.

2. The question of context raises a number of interrelated issues, including that of thematic unity, the effect of background knowledge on test results (the "hot house special" problem), the effect of topic on variation in learner language output, and the effect of "internal context" (establishing shared information and shared conventions [Brown, 1986]) on comprehension. We must come to grips with what exactly it means to speak of "contextualized" language tests. In terms of research, the question of test-retest, and intra-/inter-rater relaibility studies seems relevant here. Claire Kramsch (personal communication) has questioned whether test raters might not be "socialized" into being good raters. I would pose the question differently: What is the construct by which raters arrive at a rating? Do raters arrive at similar ratings for different reasons? We know that it is the nature of language itself that there are many ways to get the same job done. The problem we need to explore is not only the possiblity that learners with similar competence might be given different scores on our tests, but also that learners who make similar scores might do so as a result of different competence. Rater-related constructs as well as learner-related ones are of crucial importance to our understanding of variation in language production in context.

3. What specific domains of language use will the ACTFL Guidelines and tests be directed toward? We need to go beyond the relatively simple notion of "thematic unity," it seems to me, and focus attention on "purpose", or in Widdowson's (1983) terms, "procedural activity." Widdowson proposes a focus on procedures partly as a way out of a "specificity dilemma"--the logical absurdity of idiosyncratic teaching and testing materials that many see as the entailment of an LSP approach. We must consider ways of testing learners' abilities to "make sense" of linguistic input relative to specific domains of use, and answer the question, "Proficiency for what?" It may be that we can learn from the experience of the Council of Europe's Modern Languages Project, which has established "waystage" and "threshold" level specifications based upon "communicative needs" for syllabus and test development. John Trim (1984) captures an essential feature of the Council of Europe approach: "...a learning biography consists not of a straight-line progress from elementary to intermediate to advanced, but an accumulation of life-related learning experiences" (p. 20). Obviously, our thoughts on criterion- vs. domain-referencing may be relevant here, but the question of purpose in listening tests must remain a key one. This European perspective, encapsulated in the Trim quote, is quite fundamentally different from that implied in the ACTFL Guidelines, I believe. It is that language is learned socially, purposefully and cumulatively in life experience. It follows, the Council of Europe team suggests, that language should be taught and tested experientially, in life-related units. Liskin- Gasparro (1987) has referred to the difficulty of incorporating culture into the proficiency guidelines, and it has been the experience of examiners in the ILR oral interview (Pardee Lowe, personal communication) that culture only becomes an issue somewhere around Level 3. At the same time, as Clark and

Clifford (1987,p. 4) indicate, "the performance of students within most regular academic settings (including even college language majors; see Carroll, 1967) would not be expected to go beyond Level 2 or 2+." Can it be that culture need not/cannot be taken into account in our proficiency tests in academic programs?

Gass (1987) has suggested that we need to take into account the way language is learned as we develop tests and guidelines. In this context, we need to remember that language is learned in culture and in life experience. Selinker and Douglas (1985) have used the term "slices of life" to refer to the internal contexts within which, they hypothesize, interlanguages are constructed. It may be that the insistence upon viewing language learning linearly--from elementary to intermediate to advanced--distracts us from an essential fact: language learning is not linear; language is learned in units of experience. It seems to me that we cannot produce tests which represent a linear view of learning and then try to see how they are correlated to "external" criteria within life experiences; the purposeful, life-related elements must be there from the start.

4. Under the heading of technology in listening comprehension testing, I noted two issues: computer adaptive testing, and the use of interactive audio/video techniques. We need to reflect upon CAT in terms suggested by Canale (1986), who calls for studies of adaptive tests constructed outside the possibly restrictive psychometric limits of item response theory in order to determine what essential features of communicative language proficiency might be lost by adhering to those restrictions.

The question of IAV is primarily one, it seems to me, of the interface between learner and machine, and its effect on language performance--both what technology does to language and what it does to learners (e.g. Carol Chapelle, at Iowa State University, is studying the relationship between various learner characteristics and computer assisted language learning).

One might suggest (only somewhat tongue-in-cheek) that we are searching for a Grand Unification Theory in language testing that will relate constructs associated with language competence, domains of use, stages of learning, psychometric requirements and technological potential. The continuing search will be worth participating in.

References

Asher, J.J. (1969). The total physical response approach to second language learning. *Modern Language Journal, 53*, 3-17.

Auerbach, E. (1986). Competency-based ESL: One step forward or two steps back? *TESOL Quarterly, 20*, 411-429.

Bachman, L., & Savignon, S. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. *Modern Language Journal, 70*, 380-390.

Bowen, J.D. (1976). Current research on an integrative test of English grammar. *RELC Journal, 7*, 30-37.

Brown, G. (1986). Investigating listening comprehension in context. *Applied Linguistics, 7*, 284-302.

Canale, M. (1984). Considerations in the testing of reading and listening proficiency. *Foreign Language Annals, 17*, 349-357.

Canale, M. (1986). The promise and threat of computerized adaptive assessment of reading comprehension. In C. Stansfield (Ed.), *Technology and language testing* (pp. 29-46). Washington, DC: TESOL.

Canale, M., Child, J., Jones, R., Liskin-Gasparro, J., & Lowe, P. (1984). The testing of reading and listening proficiency: A synthesis. *Foreign Language Annals, !7*, 389-92.

Carroll, J.B. (1967). The foreign language attainments of language majors in the senior year: A survey conducted in U.S. colleges and universities. Cambridge, MA: Graduate School of Education, Harvard University.

Cartier, F.A. (1968). Criterion-referenced testing of language skills. *TESOL Quarterly, 2*, 27-32.

Chaudron, C., & Richards, J. (1986). The effect of discourse markers on the comprehension of lectures. *Applied Linguistics, 7*, 113-127.

Child, J.R. (1987). Language proficiency levels and the typology of texts. In H. Byrnes & M. Canale (Eds.), *Defining and developing proficiency: Guidelines, implementation and concepts* (pp. 97-104). Lincolnwood, IL: National Textbook.

Clark, J., & Clifford, R. (1987). The FSI/ILR/ACTFL proficiency scales and testing techniques: Development, current status and needed research. This volume.

Cziko, G. (1983). Psychometric and edumetric approaches to language testing. In J.W. Oller (Ed.), *Issues in language testing research* (pp. 289-307). Rowley, MA: Newbury House.

Dahl, R.C., & Luckau, P. (1985). Videodeutch: A computer assisted approach to verbal and non-verbal cultural literacy. *CALICO Journal, 2*, 13-19.

Dandonoli, P. (1987). ACTFL's current research in proficiency testing. In H. Byrnes & M. Canale (Eds.), *Defining and developing proficiency: Guidelines, implementation and concepts* (pp. 75-96). Lincolnwood, IL: National Textbook Company.

Delamere, T. (1985). Notional-functional syllabi and criterion-referenced tests: The missing link. *System, 13*, 43-47.

Denham, C.H. (1975). Criterion-referenced, domain-referenced and norm-referenced measurement: A paralax view. *Educational Technology, 15*, 9-13.

Dunkel, P. (1986). Developing listening fluency in L2: Theoretical considerations. *Modern Language Journal, 70*, 99-106.

Eastmond, J.N., & Mosenthal, R. (1985). The World Center for Computing's pilot videodisc project for French language instruction. *CALICO Journal*, *2*, 8-12.

Emmet, A. (1985). Test in English for Educational Purposes (TEEP): A new test from the Associated Examining Board (AEB). *Language Testing Update*, *1*, 3.

Faerch, C., & Kasper, G. (1986). The role of comprehension in second-language learning. *Applied Linguistics*, *7*, 257-274.

Fouly, K.A., & Cziko, G. (1985). Determining the reliability, validity and scalability of the graduated dictation test. *Language Learning*, *35*, 555-566.

Friel, M. (1984). The discourse co-operation test. *System*, *12*, 251-262.

Gale, L.E. (1983). Montevidisco: An anecdotal history of an interactive videodisc. *CALICO Journal*, *1*, 42-46.

Gary, J., & Gary, N. (1981). Caution: Talking may be dangerous to your linguistic health. *IRAL*, *19*, 1-14.

Gass, S. (1987). L2 vocabulary acquisition. This volume.

Glisan, E.W. (1985). The effect of word order on listening comprehension and pattern retention: An experiment in Spanish as a foreign language. *Language Learning*, *35*, 443-469.

Gradman, H., & Spolsky, B. (1975). Reduced redundancy testing: A progress report. In R. Jones and B. Spolsky (Eds.), *Testing language proficiency* (pp. 59-70). Arlington, VA: Center for Applied Linguistics.

Henning, G., Gary, N., & Gary, J. (1983). Listening recall: A listening comprehension test for low proficiency learners. *System*, *11*, 287-293.

Kramsch, C. (1986). From language proficiency to interactional competence. *Modern Language Journal*, *70*, 366-371.

Lantolf, J., & Frawley, W. (1985). OPT: A critical analysis. *Modern Language Journal*, *69*, 337-345.

Larson, J., & Madsen, H. (1985). Computerized adaptive language testing: Moving beyond computer-assisted testing. *CALICO Journal*, *2*, 32-36.

Little, D., & Davis, E. (1986). Interactive video for language learning: The autotutor project. *System*, *14*, 29-34.

Liskin-Gasparro, J. (1987). The ACTFL proficiency guidelines: An update. This volume.

Lowe, P. (1986). Proficiency: Panacea, framework, process? A reply to Kramsch, Schulz and particularly to Bachman and Savignon. *Modern Language Journal*, *70*, 391-397.

Madden, C. G., & Samuda, V. (1985). Task-based test design. *Language Testing Update*, *1*, 11-15.

Mydlarski, D., & Paramskas, D. (1985). Template system for second language aural comprehension. *CALICO Journal*, 3, 8-12.

Perkins, K., Brutton, S., & Angelis, P. (1986). Derivational complexity and item difficulty in a sentence repetition task. *Language Learning*, *36*, 125-141.

Powers, D.E. (1986). Academic demands related to listening skills. *Language Testing*, *3*, 1-38.

Richards, J.C. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly*, *17*, 219-240.

Rowe, A.A. (1985). Interactive language simulation systems: Technology for a national language base. *CALICO Journal*, *2*, 44-47.

Savignon, S. (1985). Evaluation of communicative competence: The ACTFL provisional proficiency guidelines. *Modern Language Journal*, *69*, 129-134.

Schneider, E.W., & Bennion, J. (1983). Veni, vidi, vici via videodisc: A simulator for instructional conversations. *System, 11*, 41-46.

Schulz, R. (1986). From achievement to proficiency through classroom instruction: Some caveats. *Modern Language Journal, 70*, 373-379.

Schrupp, D.M., Bush, M., & Mueller, G. (1983). Klavier im haus: An interactive experiment in foreign language instruction. *CALICO Journal*, 1, 17-21.

Selinker, L., & Douglas, D. (1985). Wrestling with "context" in interlanguage studies. *Applied Linguistics, 6*, 190-204.

Stansfield, C.W. (1986). Introduction to C.W. Stansfield (Ed.), *Toward communicative competence testing*. Proceedings of the Second TOEFL Invitational Conference (pp. ix-xv). TOEFL Research Report 21. Princeton: Educational Testing Service.

Tarone, E. (1983). On the variability of interlanguage systems. *Applied Linguistics, 4*, 142-163.

Tollefson, J.W. (1986). Functional competencies in the U.S. refugee program: Theoretical and practical problems. *TESOL Quarterly, 20*, 649-664.

Trim, J.L.M. (1984). Extract from Developing a unit/credit scheme of adult language learning. Reprinted in J.A. Van Ek & J.L.M. Trim (Eds.), *Across the threshold* (pp. 9-26). Oxford: Pergamon.

Tung, P. (1986). Computerized adaptive testing: Implications for language test developers. In C. W. Stansfield (Ed.), *Technology and language testing* (pp. 11-28). Washington, DC: TESOL.

Widdowson, H. (1983). Learning Purpose and Language Use. Oxford: Oxford University Press.

Woods, A., & Baker, R. (1985). Item Response Theory. *Language Testing, 2*, 119-140.

APPENDIX A. Text and Task Specifications for Reading (R) and Listening (L)

Level	Text Type	R/L Function	R/L Strategy
0/0+	Enumerative (numbers, names, street signs, isolated words/phrases)	Discern discrete elements	Recognize memorized elements
1	Orientational (simplest connected text, such as concrete descriptions)	Identify main idea	Skim, scan
2	Instructive (simple authentic text in familiar contexts and in predictable sequence)	Understand facts, literal information	Decode, classify
3	Evaluative (authentic text on unfamiliar topics)	Grasp ideas and implications	Infer, guess, hypothesize, interpret
4	Projective (all styles and forms of text for professional needs or for general public)	Deal with unpredictability and cultural references	Analyze, verify, extend hypotheses
5	Special purposes (extremely difficult and abstract text)	Equivalent to educated native reader/listener	All those used by educated native reader/listener

From: Canale, et al. (6, p. 391)

APPENDIX B

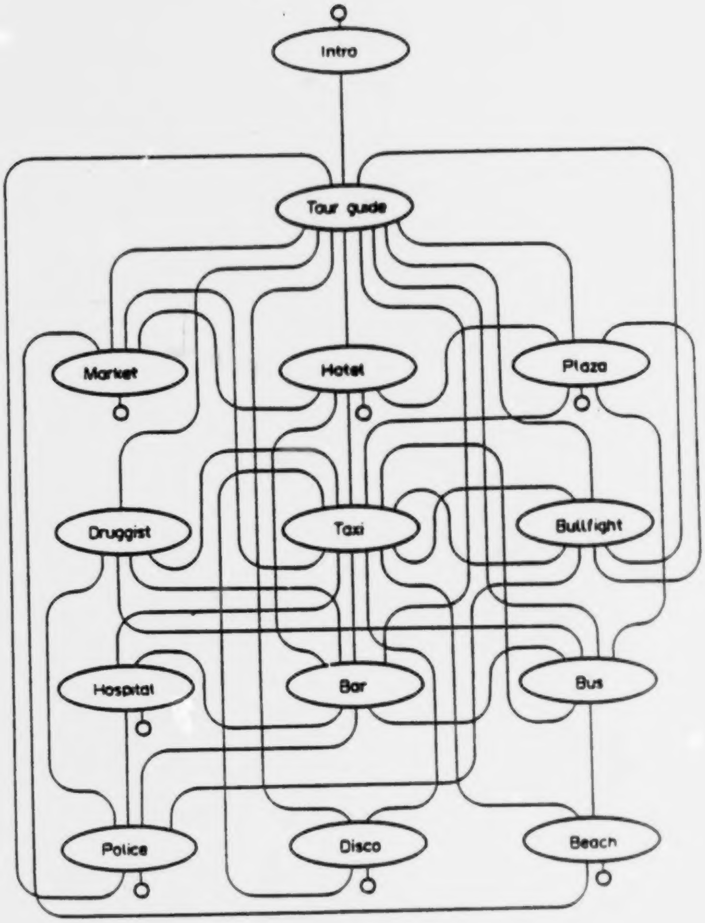

Intro
Tour guide
Market
Hotel
Plaza
Druggist
Taxi
Bullfight
Hospital
Bar
Bus
Police
Disco
Beach

Figure 2. Map of "Montevidisco" (male side).

(from Schneider and Bennion, p. 45)

DEFINING TESTS FOR PROFICIENCY IN LISTENING COMPREHENSION: A RESPONSE TO DAN DOUGLAS'S "TESTING LISTENING COMPREHENSION"

John H.A.L. de Jong
Cito, Dutch National Institute for Educational Measurement, Arnhem, Netherlands

Dan Douglas's paper is very rewarding, because it presents a comprehensive overview of recent developments in listening comprehension testing, and because it raises a number of basic issues in relation to the ACTFL Proficiency Guidelines (American Council on the Teaching of Foreign Languages, 1982, 1986).

To a large extent I agree with Douglas' views on the ACTFL Proficiency Guidelines, as also with his suggestions for development work and validation studies in listening comprehension testing. Neverthless, there are some points I would like to discuss. In my reaction I will embed my remarks in a few general observations on proficiency testing and on the nature of tests aiming to assess foreign language listening comprehension.

1. Proficiency levels

1.1 Validity of the ACTFL Proficiency Scale – In his discussion of the ACTFL Proficiency Guidelines, Douglas called attention to the insufficient empirical foundation for the Guidelines as developmental stages of proficiency. This lacking of validational studies has repeatedly been denounced (Savignon, 1985; Bachman & Savignon, 1986; Bachman, 1987) and has been the rationale for this Symposium, and I agree with Dan Douglas that, even lacking these validational studies, there is reason to question some of the principles for level distinction in the Guidelines.

Does, for example, the ability to "detect emotional overtones" (Advanced Plus) in orally presented discourse necessarily require a higher level, than the ability to "get some main ideas" (Novice-High)? This example from the "provisional" version of the Guidelines (ACTFL, 1982) seems to betray an influence of the literary tradition on the evaluation of the difficulty of oral skills, almost three quarters of a century after the publication of de Saussure's "Course". And, in the new Guidelines (ACTFL, 1986) this level of proficiency is rated even higher: "sensitivity for...affective overtones" has moved up to the Superior level. In written discourse "reading between the lines" may be argued to require a higher ability than a literal interpretation, in oral discourse emotion is often expressed at a level that requires no specific knowledge of a language. The technician who does the tape editing for Cito listening comprehension tests, though unable to understand a word of French beyond bonjour and an occasional cognate, can follow what is going on in a discussion in French at emotional and sociolinguistic levels: who is skillfull in turn taking, whose arguments are ridiculised, who is trying to overrule the others. This competence is of great functional value, because it enables him to make imperceptible cuts.

However, it is not primarily because, as Douglas stated, "it is beyond our present understanding of how language is structured", that we cannot at present evaluate the validity of the level distinction in the Guidelines. For a description of levels of language proficiency we need to know (a) how the human mind processes language, and (b) how a language is acquired. Language structure is an object of study for linguists. Educationalists developing tests to assess students' knowledge and skills may profit more from models in cognitive psychology and artificial intelligence (Habel, 1985; McClelland, Rummelhart et al., 1986; Schwab & Numbaum, 1986; Van Dijk, 1985; Van Dijk & Kintsch, 1983; Winograd, 1980, 1983).

1.2 Stages of language acquisition – The ACTFL Proficiency Guidelines purport to represent a "design for measuring and communicating foreign language proficiency": a set of criteria, constituting a variable, along which to assess language proficiency, and a set of definitions to report the outcomes of the assessment. They are becoming, however, to be regarded also as a useful instrument for curriculum planning: a hierarchically ordered set of goals to be achieved by students of a foreign language. But the scale underlying the Proficiency Guidelines clearly does not represent a set of instructional goals as is apparent from the description of the Novice-Low Level: no teacher would set as a goal to bring students to a level where they are "unable to function", or --as in the 1986 version-- where they have "essentially no ability to comprehend even short utterances."

The levels described in the Proficiency Guidelines can form a scale only if they reflect the development which actually occurs during language learning. For foreign languages that are taught in schools, this development is determined or, at least influenced, by the design of the curriculum. But, extending the notion of the Guidelines to a prescriptive function in curriculum design adds to the confusion of pedagogical and psychometric principles, and makes validation efforts circular.

A general factor, for example, is bound to be found in a number of subskills, if tests are administered to a group of individuals who have received a similar kind of training (cf. Carroll, 1961). Henning, Hudson and Turner (1985) claim undimensionality among a number of different subskills. Though the group of subjects taking the tests was very heterogeneous with respect to their first language, the authors present no reason to assume structural variation in the construct of English language abilities for the members of the group. Apparently most, if not all, had taken the TOEFL test and had passed at the level required by UCLA, thus having surpassed the elementary stage of foreign language acquisition. One can safely assume that to arrive at this level, the majority of the subjects will have undergone foreign language instruction, most probably geared at passing the TOEFL. Language instruction will have been traditional and will have covered the skills distinguished in the TOEFL and in the ESPLE test battery used by Henning et al. (1985).

Spolsky (1985b) considers that to define "what it means to know a foreign language," each of the three main approaches now in linguistic theory has to be taken into account: the "structural", the "functional" and the "General Proficiency" approach. Knowledge of a language, then, has to be described in terms of (a) knowledge of the specific elements of the autonomous linguistic system, (b) the ability to function in the language, and (c) a hypothesized general proficiency.

In the testing situation it would seem that at the very beginning stages of learning a foreign language the structuralist approach would be most appropriate. Students have been presented with a delimited number of language units. Students' knowledge at this stage is elementary in the literal sense: they know some elements. Tests need to be discrete-point, and criterion-referenced. Though administration of proficiency tests is feasible, such tests will yield no direct information with respect to a general model of developmental stages, or to "universals" in language acquisition, as knowledge at these stages is curriculum dependent.

At intermediate stages, a certain amount of functioning in the language can be learned, and, to do justice to the acquired level of language mastery, a functional approach in testing is required. The Threshold Level movement, initiated by the Council of Europe (Van Ek, 1975), and the Graded Objectives movement in the United Kingdom (cf. Southern Examining Group, 1986) acknowledge the curriculum-bound aspect of their respective approaches by spelling out complete lists of notions and functions to be mastered at different stages of functional ability. The claim of the ACTFL Proficiency Guidelines to represent "a design for measuring and communicating foreign language proficiency" regardless of the curriculum overlooks the possibility of variation in sequencing the acquisition of distinct functions, and is possibly explained by the FSI origin, where the instrument was used in testing for higher stages of functional language proficiency.

At higher stages, where proficiency approaches normal--native speaker like--functioning, subjects' control of a language suffers less from occasionally occurring deficiencies. The language learner has acquired enough versatility in the language to circumvent such inconveniences. Competence to express one's needs without the knowledge of the appropriate language units is an elementary and essentially non-linguistic competence. At these stages, however, this strategic competence enables an effective use of acquired language elements, thereby bringing language use at a more native-like level. At these stages only, subjects can become independent of the curriculum, their proficiency can be assessed on a single dimension and the General Language Proficiency approach is more likely to apply than at earlier stages of acquisition.

2. Testing proficiency levels

2.1 Instructional goals – Tests constructed to assess students' proficiency should be distinguished from tests designed to be used in instruction. There is no doubt that teaching and testing are related in an intricate way. A clear definition of instructional goals is a prerequisite for the construction of equitable tests, and defining levels of proficiency for a testing purpose can lead to a better definition of instructional goals.

Components of the language learning process, however, are not necessarily appropriate goals in proficiency testing. An essential element in language learning, e.g., is the acquisition of an extensive vocabulary, but testing vocabulary items is irrelevant from the proficiency perspective. And, while learning lexical elements and acquiring full understanding of where and when to use them may constitute consecutive steps in language learning, no one in his right communicative mind will think of testing these steps separately.

Power's (1986) survey of academic demands related to listening comprehension reveals another aspect of the teaching/testing dichotomy. Among the most important subskills he mentions are "identifying major themes or ideas" and "identifying the topic of a lecture." One will easily agree that indeed for foreign students studying at American universities these are essential skills to be able to function in academia, and therefore should be a goal of their learning of English. For the language tester, however, the problem is that these skills are apparently rather easy for the non-native students. In fact "identifying the topic" is about as easy as "recognizing the function of non-verbal signals." Foreign students do have great difficulties with "retaining information through note taking" and "retrieving information from notes." But the language tester will find it hard to design effective, discriminating tests for these skills, as they are also the hardest to master for native speakers. The measurement of these skills would therefore deviate substantially from the dimension of foreign language proficiency.

2.2 Test objectives – Validity and reliability of listening comprehension tests would be served best by setting tasks that (a) are relevant and (b) discriminate maximally between able listeners (native speakers) and those wishing to achieve that ability (non-natives). From Powers's (1986) report, the skills that fit both these requirements best seem to be "recognizing the function of structure words," "comprehending key vocabulary," "deducing the meaning of words from context," and "inferring relationships between information."

The central role of lexical knowledge in listening comprehension--for natives and non-natives--has also been shown in empirical studies (Clark, 1973; Groot, 1975, 1977), and is probably explained by the fact that lexical knowledge is essential for a correct processing of orally presented discourse, already at the phonological level (Matter, 1986; Koster, 1987).

Apart from subskills operating mainly at the lexical level, an element of speed would contribute considerably to the discriminating power of listening comprehension tests (Powers, 1986). In an experiment with word recognition tasks, Koster (1987) found that under time pressure non-natives tend to disregard contextual constraints, though given time they make more use of context than native speakers do. Clark (1973) reports that almost ten per cent of the variance on the test battery he used could be attributed to fluency or speed in associating words.

These criteria would agree largely with recommendation No. 4 quoted by Dan Douglas from Canale et al. (1984): "scoring criteria should be accuracy of comprehension, independence in dealing with text, and speed of completion" and bring us back to over 25 years ago when Carroll (1961) required that listening comprehension tests measure "rate and accuracy of listening."

2.3 Tasks in listening comprehension tests – Douglas offers the following categories of listening comprehension tests:

1- Listen and Get the Right Answer Tasks;
2- Reduced Redundancy Tasks;
3- Repetition/Imitation Tasks;
4- Interaction Tasks;
5- Media-Transfer Tasks.

He admits that the categories are fuzzy. I would add that they tend to confuse different classes of task setting and focus primarily on the response demanded from the test taker.

A categorization of listening comprehension tests including predominantly categories of response tasks is too restricted from a proficiency perspective. In a listening comprehension test three classes of tasks can be distinguished: listening tasks, comprehension tasks, and response tasks.

2.4 Listening tasks – The samples of language used in a test constitute the first class of tasks. The samples are selected by the test developer from all possible instances of orally presented discourse and from the material to which the examinee will have to listen. An example of a listening task could be: "Listen to the following sixty seconds of tape recorded speech taken from the lecture of a visiting professor speaking at a rate of 200 wpm, on the effects of alcohol". The difficulty of a listening task is determined mainly by acoustic and linguistic factors such as amount of background noise, voice quality, speed of delivery, vocabulary, and abstraction level of the topic. Different listening tasks can be set by varying these factors.

2.5 Comprehension task – The questions in a test operationalize the second class of tasks. The comprehension task accompanying the listening task cited above may be: "What is the main idea presented in this passage from the lecture," or "What did the speaker say about detail X?"

The difficulty of the comprehension task is determined by the difficulty of the listening task, by the comprehension task itself, and by the relation between both tasks. Given the same listening task, the difficulty of a test item can be varied by setting different comprehension tasks. A question directed at the "main idea" can sometimes be easier than a detail question, like when the main idea is expressed more or less literally and redundantly, in plain and simple words.

2.6 Response tasks – The reactions demanded from the test taker constitute the third class of tasks. The response task may be operationalized by e.g.: "give a description of all details concerning aspect X," or "recognize which of the four statements printed in your test book is the best paraphrase of the main idea."

The difficulty of the response task is determined by the listening task, the comprehension task, the response task itself (the type of response requested, the phrasing of the alternatives in the case of a multiple choice question or the demands on productive skills in open-ended formats), and by the relations between all three tasks. Given the same listening and comprehension tasks, item difficulty can be varied by setting different response tasks.

Most of the error in testing listening comprehension is due to the operationalisation of this class of tasks. In open-ended questions a productive ability influences the response; in multiple choice questions the comprehension of the alternatives may require a specific type of ability. Unfortunately, the response tasks are the only class of tasks demanding overt behavior, and it is on account of performance of these tasks that listening comprehension ability is assessed.

2.7 Test description – Within the test situation, which is perforce artificial (cf. Spolsky, 1985a), the listening task is the most open task and is closest to "normal" language proficiency, whereas the response task is most heavily constrained and represents the most artificial level. The assessment of listening comprehension, however, is based only on the student's performance of the response task.

A major problem in comprehension testing is the discrepancy between the above mentioned classes of tasks. The amount of discrepancy between these classes adds up to lack of validity.

To control this discrepancy, the development of tests, and subsequently their description too, should start out from a definition of the listening task: "what type of language use should the student be able to listen to?" The next step is to define the comprehension task(s): "what should the student be able to understand from this particular type of language use?" These are essentially linguistic issues. The response task involves also linguistically irrelevant factors, even down to the time available for administering and scoring the test.

The Guidelines offer a description of a number of levels in listening comprehension. The description pertains mostly to listening tasks. Comprehension tasks are referred to at the lower levels, but are not described as communicatively relevant functions: "understanding isolated words," "understanding short learned utterances," "understanding sentence length utterances". Response tasks are seldom mentioned, and their description is either negative ("requires long pauses," "may require repetition"), or unclear ("shows awareness of ...," "shows sensitivity to..."). The Guidelines offer little specifity as to the constituent subskills of listening comprehension ability (cf. Richards, 1983; Powers, 1986).

2.8 Test dimensionality – Ever since some of the basic assumptions underlying measurement in general have been judged to be consistent with principles of educational measurement and have been formulated explicitly in Item Response Theory, educationalists have started worrying about assumptions such as local stochastic independence and unidimensionality. Dan Douglas referred to Dandonoli (1987), Canale (1986) and Tung (1986) to illustrate these concerns.

First, the assumptions of local independence and unidimensionality are implicitly accepted in the framework of classical test theory.

Second, unidimensionality in the definition of Item Response Theory does not necessarily imply an undivisable trait: it does, however, imply a fixed relation between constituent subskills.

Third, if the unidimensionality assumption cannot be met because of the complexity of the skill, or does not apply as in the case of sequentially arranged units of instruction, techniques to select unidimensional subsets of items can be applied (de Jong, 1987; Gustafson, 1980), or else multidimensional models can be used (Heartel, 1985, McKinley & Reckase, 1982, 1984; Reckase, 1985; Verhelst, 1982; Verhelst & Glas, 1984).

Fourth, if the unidimensionality assumption cannot be met, the report of the outcomes of the assessment will have to take account of pluridimensionality, components will have to be scored separately, and a profiled report will have to be produced. One cannot at the same time call for integrated and integrative tests, question the assumption of unidimensionality and continue expressing the evaluation of a student's ability with a single mark, be it A, B, C, or 1, 2, 3.

3. Concluding remarks

The Proficiency Guidelines can become a valid measurement design if language teachers formulate their goals in parallel to these Guidelines. In that case, however, the Guidelines will no longer constitute a design for a "global assessment" (ACTFL, 1986) of proficiency, but for an accurate measurement of achievement.

The issue then is not whether the Guidelines form a scale, but whether the curriculum designed to conform to these Guidelines constitutes an acceptable set of goals ordered in such a way as to be justifiable by pedagogical and/or economic principles.

The decision in a number of European countries to define the minimum requirements for foreign language instruction in terms of a survival level, and to set the survival level as an instructional goal, was a political--not a linguistic--issue.

Developing the Guidelines as "a design for measuring" is like working backwards, designing "how" to measure, before defining "what" to teach. Working backwards may have produced a considerable amount of highly relevant discussions and a better understanding of what language teaching and testing should be about. In order to decide exactly "how" one should test, one has to start at the beginning, and decide "what" should be taught.

References

American Council on the Teaching of Foreign Languages. (1982). *ACTFL provisional proficiency guidelines*. Hastings-on-Hudson NY: Author.

American Council on the Teaching of Foreign Languages. (1986). *ACTFL proficiency guidelines*. Hastings-on-Hudson NY: Author.

Bachman, L. F. (1987). *Problems in examining the validity of the ACTFL oral proficiency interview*. This volume.

Bachman, L. F., & Savignon, S. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. *Modern Language Journal, 70*, 380-390.

Canale, M. et al. (1984). The testing of reading and listening proficiency. *Foreign Language Annals, 17*, 389-392.

Canale, M. (1986). The promise and threat of computerized adaptive assessment of reading comprehension. In C. W. Stansfield (Ed.), *Technology and language testing* (pp. 29-46). Washington DC: TESOL.

Carroll, J. B. (1961). Fundamental considerations in testing for English language proficiency of foreign students. In *Testing the English proficiency of foreign students*. Washington DC: Center for Applied Linguistics.

Clark, M. L. (1973). *Hierarchical structure of comprehension skills: Vol. 2. Factorial and "smallest space" analysis of primary reading and listening correlations: An empirical study of Grade 7 children's performance* (ACER Research Series No. 49). Hawthorn, Victoria: Australian Council for Educational Research.

Dandonoli, P. (1987). ACTFL's current research in proficiency testing. In H. Byrnes & M. Canale (Eds.), *Defining and developing proficiency: Guidelines, implementation and concepts* (pp. 75-96). Lincolnwood IL: National Textbook Company.

de Jong, J. H. A. L. (1983). Focusing in on a latent trait: Construct validation by means of the Rasch model. In J. van Weeren (Ed.), *Practice and problems in language testing 5: Non-classical test theory; Final examinations in schools* (pp. 11-35). Arnhem: Cito, National Institute for Educational Measurement.

de Jong, J. H. A. L., & Glas, C. A. W. (1987, April). *Validation of listening comprehension tests using Item Response Theory*. Paper presented at the Colloquium on Language Testing Research: Priorities and Prospects, Reading, MA.

Groot, P. J. M. (1975). Testing communicative competence in listening comprehension. In R. L. Jones & B. Spolsky (Eds.), *Testing language proficiency*. Arlington, VA: Center for Applied Linguistics.

Groot, P. J. M. (1977). *Luistervaardigheid Frans, Duits, Engels: Doelstelling en toetsing*. [Listening comprehension of French, German, English: Objectives and testing]. Amsterdam: Meulenhof.

Gustafson, J. E. (1980). Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology, 33*, 205-233.

Habel, C. U. (1985). Referential nets as knowledge structures: Some structural and dynamic properties. In T. T. Balmer (Ed.), *Linguistic dynamics: Discourse procedures and evolution*. Berlin: Walter de Gruyter.

Heartel, E. (April, 1985). *A comparison of latent structure models for reading comprehension items*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.

Henning, G., Hudson, T., & Turner, J. (1985). Item Response Theory and the assumption of unidimensionality for language tests. *Language Testing, 2*, 141-154.

Koster, C. J. (1987). *Word recognition in foreign and native language: Effects of context and assimilation*. Dordrecht: Foris Publications.

Matter, J. F. (1986). *A la recherche des frontières perdues: Étude sur la perception de la parole en français* [In search of the lost boundaries: Word recognition in spoken French]. Amsterdam: Editions de Werelt.

McClelland, J. L., & Rummelhart, De. E. & the POP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

McKinley, R. L., & Reckase, M. D. (1982). *The use of the general Rasch model with multidimensional item response data* (Research report ONR 82-1). Iowa City, IA: The American College Testing Program.

McKinley, R. L., & Reckase, M. D. (1984). *A latent trait model for sequentially arranged units of instruction* (Research report ONR 84-2). Iowa City, IA: The American College Testing Program.

Powers, D. E. (1986). Academic demands related to listening skills. *Language Testing, 3*, 1-38.

Reckase, M. D. (1985, April). *The difficulty of test items that measure more than one ability*. Paper presented at the meeting of the American Education Research Association. Chicago.

Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly, 17*, 219-240.

Savignon, S. (1985). Evaluation of communicative competence: The ACTFL provisional proficiency guidelines. *Modern Language Journal, 69*, 129-134.

Schwab, E. C., & Nusbaum, H. C. (Eds.). (1986). *Pattern recognition by humans and by machines. Volume 1: Speech perception*. London: Academic Press.

Southern Examining Group (1986). *General certificate of secondary education: French, 1988 examinations*. Oxford: The Southern Examining Group.

Spolsky, B. (1985a). The limits of authenticity in language testing. *Language Testing, 2*, 31-40.

Spolsky, B. (1985b). What does it mean to know how to use a language? *Language Testing, 2*, 180-191.

Tung, P. (1986). Computerized adaptive testing: Implications for test developers. In C. W. Stansfield (Ed.), *Technology and language testing* (pp. 11-28). Washington, DC: TESOL.

Van Ek, J. A. (1975). *The threshold level*. Strasbourg: The Council of Europe.

Van Dijk, T. A. (1985). Strategic discourse comprehension. In T. T. Balmer (Ed.), *Linguistic dynamics: Discourse procedures and evolution*. Berlin: Walter de Gruyter.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. London: Academic Press.

Verhelst, N. D. (1982). *Latente eigenschap modellen en multidimensionele representatie*. [Latent trait models and multidimensional representation.] Utrecht: State University, PSM Progress Reports.

Verhelst, N. D., & Glas, C. A. W. (1984). *Using the Rasch model in the analysis of multifactorial (quasi) experiments and interrupted time series*. Utrecht: State

University, PSM Progress Reports.

Winograd, T. (1980). What does it mean to understand language? *Cognitive Science, 4*, 209-241.

Winograd, T. (1983). *Language as a cognitive process, Volume I: Syntax*. Reading, MA: Addison-Wesley.

TESTING LISTENING COMPREHENSION: A RESPONSE TO DAN DOUGLAS

Harry L. Gradman
Indiana University

For a variety of reasons, I enjoyed reading this paper. I found myself in agreement much of the time, and I liked the straight-forward presentation in a set of categories that made some sense to me. The Douglas paper also helped to calm my nerves just a bit, because the more I thought about listening comprehension, especially in relation to the ACTFL Oral Interview, after Professor Valdman first mentioned this session, the more distressed I became. I had always had a hard time separating listening comprehension from speaking as totally separate skills in normal linguistic behavior, quite distinct from my ability to separate reading and writing, though they too were a receptive-productive skills pair, just like listening and speaking. What I discovered in Dan Douglas's comments was that his focus had the seeds of some of my own thoughts and doubts which I want to talk about under the headings of isolation of listening comprehension or "purity" or "contamination", context, efficiency, guideline specificity (or vagueness), and needs for broader evaluative data.

When I asked some colleagues of mine the question "What is listening comprehension?" their answers prompted a succession of smiles at their inability to talk about listening comprehension without mentioning a broad range of other factors as well. They seemed to be perfect examples of the fact that we can't separate the person who speaks the language from the knowledge that the language gets at--call it "intelligence" or "knowledge of the world" or--"experience" or whatever. How do we isolate listening comprehension from non-basic-linguistic knowledge? How do we keep it "pure" listening comprehension testing short of Asher-type techniques of "stand up, sit down" or early Lado style very discrete minimal pair type questions ("ship-sheep")? And when we get beyond the question of listening comprehension and knowledge or experience, we can't separate it from the other skills of speaking, reading, or writing. Normally, we find out about listening comprehension through a person's oral performance, or reading, as in the case of tests requiring a person to read and select the appropriate choice after some spoken message. In some cases, passages are dictated, with responses written. I have no quarrel with the fact that listening comprehension is involved, but I don't know what the weight is of the other factors. I don't know what or how much to separate out as listening comprehension and what is something else. Other factors associated with oral interviews, such as the dynamics of the interview itself, the range of personality factors that enter into normal communication, serve only to compound the issue. I just don't see how to talk about listening comprehension in isolation, at least with regards to this test. And the minute I can't isolate it, I wonder how much focus to give it as a consideration in testing.

"Context" is a compounding issue. Just exactly how much "context" is enough? As aware as we are of some positive features of more extended context, just how much context do we need? Is much necessary? Can context actually get us into trouble by moving beyond the more bland attempt to picture language, say at the discrete sentence level, and touch on areas that may not be "characteristic" of a large segment of language use? We can get false pictures because of a person's ability to perform in some frameworks and not in others. When we first gave the noise test in contextualized form here, what we discovered was not that we got different results--different proficiency ratings--but rather that the students found the test more enjoyable--if that can ever be the case. They felt better taking the test, an admirable feature, of course, but not necessarily more efficient than the discrete sentence version we had used before. It may be true that normal language use has us using language in context, but the language ability of native-like speakers allows them to respond to and deal with sentences in isolation.

An important consideration for us all is that of practical efficiency. Any oral interview procedure is labor intensive, at a time when resources for teaching, much less testing, are simply in short supply. I'm not so worried about the reliability of the test as I am about the cost of the test compared to other methods that might produce similar results. This, of course, is an area that deserves continued study--something I'd like to see done more in the ESL area, because the situation in this country is in some ways different from many other foreign language teaching areas, and the implications could be different. ESL students typically have much more contact with the target language, both in their language programs themselves, many of which are intensive and require up to five or six contact hours of instruction daily, and in the daily exposure one living in the target language culture actually receives. I'm also worried in a practical sense with some of the more technological advances, at least at present. I'd hate to be so dependent on advanced technology that my ultimate test could be afforded by only a few. And yet I look at things like speech synthesizers as real potential boons as their development improves. In our preliminary tests here of synthesized sentences, we found that comprehension followed a pattern quite similar to that of the noise test, which followed a pattern quite similar to dictation. Perhaps what we should do is try dictating a contextualized paragraph and marking only the last part of it. It might be nice to see how performance on something like that correlated with the oral interview.

I don't want to spend a long time on the vagueness of the Guidelines, because I think that is sort of unfair if I can't be more precise in my own ability to isolate listening comprehension, either from knowledge, other skills, or context. I simply give the following, taken from the 1986 ACTFL Proficiency Guidelines, as exemplifying some of my concern:

> Intermediate-Low
> Able to understand sentence-length utterances which consist of recombinations of learned elements in a limited number of content areas, particularly if strongly supported by the situational context... Understanding is often uneven; repetition and rewording may be necessary. Misunderstandings in both main ideas and details arise frequently.

It's not that I don't have an idea of what this means, because, in fact, I do. I simply can't reduce my understanding to concrete terms that I can mark on a chart. I'm curious if there is a measure for how often it takes "often" to be uneven, and how uneven "uneven" has to be at the Intermediate-Low level. Sally Magnan reminded us yesterday that we live in a world of uncertainty--a world that turns on uncertainty; and I certainly can't disagree with that. I think it can be argued though that uncertainties can be lessened, and we're better off for that lessening.

Finally, I should add that our recent interest in language program evaluation in its broadest sense has led us to the concern that perhaps test scores or ratings alone aren't enough: that it's important in evaluating the language proficiency of a student to understand as well the kind of a program that person has been through, in terms of staff, facilities, and experiences the student has been exposed to. How much weight this has we don't know. But there is a tendency to think that maybe we need more description at this point before judgments made from quick analyses of partial structures should be applied.

Having said all this, I want to say some positive things about the ACTFL Oral Proficiency test from my limited experience with it, though not such limited experience with other areas of aural/oral proficiency testing. I found the interview techniques to be learnable, sensible, and relatively easy to put into practice. I found it to be an interesting--if tiring--way of gaining what I usually felt comfortable with as a fairly correct assessment of someone's language proficiency (but then again, I tested in English, my native language, and the language the test subjects were living in and, by and large, studying in). I think of it as a good way to test, though perhaps less broad in its ramifications than might be currently hoped for, at least from a testing point of view. It seems apparent that we still need more basic language research before being able to understand with depth what it is that this test taps or does. While it is fine that it works, so too do many others.

References

ACTFL Proficiency Guidelines. (1986). Hastings-on-Hudson, NY: Author.

Asher, J. J. (1969). The total physical response approach to second language learning. *Modern Language Journal, 53*, 3-17.

Educational Testing Service. (1982). Oral proficiency testing manual. Princeton, NJ: Author.

Gaies, S., Gradman, H., & Spolsky, B. (1977). Toward the measurement of functional proficiency: Contextualization of the noise test. *TESOL Quarterly, 11*, 51-57.

Gradman, H., & Spolsky, B. (1975). Reduced redundancy testing: A progress report. In R. Jones & B. Spolsky (Eds.), *Testing language proficiency* (pp. 59-70). Arlington, VA: Center for Applied Linguistics.

Greene, B., Pisoni, D., & Gradman, H. (1985). Perception of synthetic speech by nonnative speakers of English. *Research on Speech Perception, 11*, 1-10.

Lado, R. (1961). *Language testing: The construction and use of language tests, a teacher's book.* New York: McGraw-Hill.

Summary of the Discussion Session on TESTING LISTENING COMPREHENSION

Cynthia A. Fox
Indiana University

There seemed to be tacit agreement among the participants that listening comprehension is, indeed, the "quiet corner" in both the testing and the teaching fields. As a result, the first part of the discussion centered on various aspects of listening comprehension that need to be looked at in order not only to better test this skill, but also to better define it.

The discussion opened with one participant declaring that the notion of context needs expansion from a factual framework to a communicative one. It was suggested that we should be testing for the extent to which the testee can enter into the tester's frame of reference, and that the Guidelines do not sufficiently address this type of phenomenon. Another participant questioned the role of vocabulary as outlined by De Jong. What, it was asked, would a test that focused on individual items look like? How are we to differentiate between memory and comprehension? What are we to do with the problem of vocabulary in isolation versus vocabulary in context? Since aural and oral skills combine to form a two-way street, there is no such thing as conversation without comprehension--what is the interviewer doing to get the interviewee to understand? Another participant challenged the centrality of vocabulary. The idea expressed was that the work done on English, Spanish and French has biased our viewpoint, but that since there are both more-inflected and less-inflected languages, there is no basis for assuming that vocabulary is on the whole more important than syntactic or morphological indicators.

The discussion then turned to a clarification of the background of the Listening Guidelines. It was explained that the government used job-specific tests and that the experience gained from them was not factored into the proficiency test in the way those adapting the Guidelines would have hoped. Now, although a general moratorium on rewriting the Guidelines has been declared, it is agreed that the next set of definitions that must be addressed are those that deal with listening. Moreover, a tentative agenda has been set from which to attack the definitions; this is a tri-partite division consisting of interactive listening (as in the oral interview, where the listener may ask for clarification, etc.), overheard interactive listening, where the listener hears a conversation but may not enter into it, and non-participative listening, such as what one hears on the radio or on television.

Concern was then expressed by another participant for the importance of variety in testing. It was pointed out that listeners can fail to understand for many reasons, one of which might be parsing, another of which might be that the question itself is "stupid." In other words, interpreting the intention of a question is often a problem. Learners, it was further suggested, need also be taught how to monitor the nature of their failure to understand.

Following up on the earlier discussion of the background of the Guidelines and addressing a number of the issues raised thus far, another participant reported on the test the FSI currently administers which, for the past five years, has departed markedly from the OPI in that it is more task-oriented and not a single-item test. The participant then described three components of the test--a warm-up, a section of interactive listening in which the examinee interviews a native speaker on a specific topic and then reports in English on the content of the interview to the examiner, and finally a briefing in the form of a monologue on the part of the examinee. The five-factor scoring for this test includes listening comprehension, discourse competence, grammatical accuracy, fluency and lexicalisation. Within the curriculum, students are taught conversation management skills such as repairs and requests for clarification. Speaking to the question of memory, the participant's experience suggests that it is part of the examiner's role to distinguish between memory and comprehension and that, in general, this type of test depends heavily on tester technique to get a complete sample and as such the test is very labor intensive.

The next participant returned to the issue of problems with the current set of Guidelines and expressed pleasure with Dan Douglas' use of the terms integrative and integrated in describing recent developments in testing. If guidelines are to be put forth that make claims about the development of communicative ability, the participant went on to say, they should reflect this view of how language develops through life experience. Rather than talking about four skills and a disembodied notion of culture, then, we might better consider modalities, channels, interpretation, expression, and negotiation of meaning in both oral and written channels. This might lead to a better framework from which to develop an approach to testing.

Similar issues were addressed by the next participant, who reported on research that suggests that register is also a factor in comprehension. Studies show, for example, that when faced with understanding a lecture, a news item and a dialogue all on the same topic and using the same vocabulary, the same student makes very different scores. The participant then went on to say that the question of memory may be one of focus; studies also show that a student's performance is better when s/he knows in advance what to focus on. Focusing may therefore solve the problem of memory and, by reducing the content to language specific kinds of things, solve the problem of "stupid" questions at the same time.

The focus of the discussion then shifted to more general issues that had come up in this and previous sessions, beginning with the definition of criterion-referenced testing. One participant felt it was important to point out that this type of testing is not in opposition to domain-referenced testing, but in fact is basically the same thing. What makes a test criterion-referenced, the participant went on to say, is not that a cut-off score is set, but rather that the domain itself is viewed as a criterion of ability. The notion of cut-off score is an added feature that may or may not be used; a criterion-referenced test intended for diagnostic purposes would not use cut-off scores. The participant also stated that the discussion pointed to two major areas that need to be looked at, not just in terms of listening comprehension, but more generally: language for specific purposes, and the role of context, particularly in the oral interview itself.

Another general point that was raised about testing concerned the question of linearity. Linear proficiency testing, one participant claimed, is only possible at the higher end of the scales, that is, when the learner has reached the point where s/he has learned to avoid difficulties and begun to develop coping strategies to compensate for them. Prior to this stage, i.e., at the lower end of the scales, linear proficiency testing is only possible where there are uniform instructional programs. At these lower levels, then, linearity must be identified with achievement rather than proficiency testing. Another participant wondered how to fill the gap between the two stages of learning, but no one offered any suggestions.

The discussion proceeded to still other general concerns, one of them the issue of test validity. The participant who brought this up had always understood that proficiency tests were not intended as instruments for diagnostic purposes; since the OPI is global in nature, global types of feedback seem perfectly appropriate. A second concern was with curriculum planning, and the widely-held suspicion that people are planning curricula around the test. The participant suggested that if we were to distinguish acquirers from learners, we might clear some of the confusion surrounding both diagnostic testing and curriculum planning. Lastly, the participant warned the audience that his experience learning Japanese taught him that there is no such thing as sociolinguistically neutral speech, and that the issue should not be relegated to higher stages of proficiency.

The next participant responded first by heartily agreeing that both sociolinguistic and discourse competence are present at the earliest stages of language aquisition. The participant strongly disagreed, however that a learner-acquirer distinction was being made in the Guidelines. Although the claim is made that the scales represent "stages of proficiency regardless of when, where, and how the language is acquired," the participant went on to say, other statements in the Guidelines seem to promote a thinly disguised ALM methodology that has little to do with natural second language processes. By way of illustration, the participant quoted from the Guidelines' description of an early level of listening comprehension--"can comprehend complete sentences that represent recombinations of learned utterances."

The final contributor to the discussion moved the conversation back to more concrete questions of testing listening comprehension, tying this problem in with the earlier session on less commonly taught languages. The point was made that the Guidelines do not address the problem of dialect variation which, with regard to non-standardized languages, is much broader in range. Somehow, the ability of the listener to move within this broad range of normal, interactive communication must be taught and assessed.

The session came to a close after having covered a broad range of topics relating to both testing and teaching listening comprehension, as well as the place of listening comprehension within the general history of the Guidelines. The discussion also added to the on-going debates about criterion-referenced testing, linearity, test validity, and the role of sociolinguistics within the OPI. Further research into listening comprehension was called for, with respect in particular to the role of context, vocabulary, conversation management, task-oriented language, register, and dialect variation.

THE PROBLEM OF THE TARGET MODEL IN PROFICIENCY-ORIENTED FOREIGN LANGUAGE INSTRUCTION

Albert Valdman
Indiana University

1. Introduction

Learning to communicate in a foreign language is a perilous adventure in which the voyager must navigate a straight course between the Scylla of grammatical inaccuracy and the Charybdis of functional ineffectiveness. The introduction of the notion of communicative competence has led second language teaching methodologists to give higher priority to semantic appropriateness and functional adequacy. For example, in defining the Threshold Level, Van Ek (1976) placed emphasis on getting meaning across in communicative interactions and downgraded the importance of linguistic well-formedness. In the ensuing years hasty and incautious adaptations of the notional-functional syllabus to the American foreign language teaching situation have led language teachers to accept grammatically inaccurate production on the part of learners provided that it got the message across. More recently, however, more balanced views have prevailed which recognize the inseparability of grammatical accuracy, semantic interpretability, functional adequacy, discursive coherence and cohesion, and sociolinguistic appropriateness.

Within the framework of communicative competence de-emphasizing any of these competencies is reductionist. In their seminal analysis of the components of communicative competence Canale and Swain (1980, p. 5) stress the balance that must exist between linguistic and sociolinguistic competence:

> Just as Hymes (1972) was able to say that there are rules of grammar that would be useless without rules of language, so we feel that there are rules of language use that would be useless without rules of grammar.

To illustrate this point they show that, given the partially shared social and cultural matrix, sociolinguistic competence in Canadian English transfers to Canadian French but that without a minimal level of grammatical competence no effective communication with monolingual French speakers is possible. The tension resulting from the conflicting pressures of grammatical and pragmatic constraints is inherent in the use of language but it must be resolved by accommodation between them, not by the elimination of one or the other. However, while it is essential, grammatical accuracy is not sufficient for the effective use of language.

The fundamental issue in developing tests of communicative ability and in devising guidelines for the imparting of communicative skills is not the central versus the peripheral nature of grammatical accuracy or the legitimacy of a grammatically-oriented syllabus but rather the identification of target norms for grammatical well-formedness and the specification of minimum levels of linguistic competence (Canale & Swain, 1980, p. 27):

BEST COPY AVAILABLE

> It is particularly important to base a communicative approach on the varieties of the second language that the learner is most likely to be in contact with in a genuine situation, and on the minimum levels of grammatical and sociolinguistic competence that native speakers expect of second language learners in such a situation and that the majority of second language learners may be expected to attain.

The ACTFL Guidelines (in both the provisional and the 1986 versions) have set speech typical of Educated Native Speakers (ENS) of the target language as the implicit target for foreign language learners. As the Guidelines recognize, ENS[1] are capable of perceiving and producing other varieties of the language. What does not appear clearly in the Guidelines is that, in addition to shifting to other norms, ENS show highly variable usage. What we call ENS speech is in fact an idealization from highly variable use. These target users control to differing degrees a repertoire of grammatical forms, discourse styles, and sociolinguistic rules which they apply variably depending on the situational context. It is control of this repertoire, not of a particular norm, that should constitute the ultimate goal for FL learning.

In this paper I will show that the ENS target chosen by the 1986 Guidelines eliminates from consideration a range of native speech which may constitute more appropriate targets for certain learners, particularly at lower proficiency levels. Then, I will develop a model of interlinguistic variation that legitimizes non-standard varieties of the target language, and I will argue that, given the special nature of the FL teaching situation, the traditional standard norm should be replaced by dynamic pedagogical norms based not only on ENS usage but on other norms of the target language speech community. I will conclude by showing that this notion of pedagogical norm makes possible at lower proficiency levels a higher degree of grammatical accuracy by providing a better match between the target and learners' capabilities.

2. The ENS as target model

In the Guidelines' framework, the ENS constitutes the ultimate target for second-language learning. According to the Guidelines the chief difference between ENS and other speakers who are not such worthy models for foreign learners resides in ENS' more extensive repertoire. Its members have the capability of perceiving and producing features that characterize other varieties of the language. For example, the characterization of speaking ability at the Superior Level specifies that learners are "partially familiar with regional or other dialectical (sic) variants." In the area of comprehension, upper levels of proficiency are specified in terms of the ability to "understand all forms and styles of speech pertinent to personal, social, and professional needs tailored to different audiences." It is pointed out, however, that even at the Distinguished Level second-language learners "may have difficulty with some dialect and slang." Implicit in the sociolinguistic model that underlies the 1986 Guidelines is the notion that ENS are able to shift in and out of their habitual speech styles to produce speech varieties associated with social groups located below them in the social scale or in geographical areas outside of the prestige centers. Members of other social groups or residents of other geographical areas of TL countries fail to show significant stylistic manoeuver or the ability to understand or produce the speech of other groups in the overall speech community.

The notion of ENS as more skillful and deft users of the language also appears in the only explicit discussion of the target norm within the Guidelines' framework (Lowe, 1985). For Lowe, speakers of a language operate at the S-3 level in most of their daily communicative interactions; only occasionally do they move into higher levels. However, Lowe also correctly links ENS status to formal educational training and, especially, social class. He points out that the concept of ENS applies more readily to "stratified" societies such as those of France, Germany, and Great Britain, than those, such as that of the United States, that have undergone social leveling.

The belief that educated middle class speakers possess a richer speech repertoire than members of other social groups underlies currently available sociolinguistic models. In Labov's early sociolinguistic model (1964) speakers fall along a unidirectional variability continuum linking vernacular and standard poles. Individual speakers are placed along the continuum according to the relative proportion of standard variants they show for any particular variable. For example, for the θ variable in New York City English, realized as [θ], [tθ], or [θ], lower and working class speakers show a relatively flat curve as they shift among speech styles, whereas middle class speakers, notably those from the lower middle class, demonstrate a significant increase in the frequency of the valorized standard variant [θ] in more formal styles. Not surprisingly, in view of the bias in favor of ENS underlying the model, empirical studies are often designed so as to confirm the lack of variability in the speech behavior of lower class speakers (Lindenfeld, 1972; Behnsted, 1973). Lindenfeld found that French speakers from the lowest of the two social classes she had identified showed the same level of syntactic complexity in formal and informal style. But she induced so-called formal speech by asking informants to imagine they presented a lecture before an auditorium audience of a hundred persons. This is hardly a situation that is within the realm of experience of working class speakers, and it provides little information about syntactic features that they actually produce in genuine formal interactions characteristic of their social group. Behnsted designed his study of variation in French interrogative constructions on the assumption that middle class speakers showed different syntactic structures in formal and informal style whereas working class compatriots operated with a single style, see Table 1 (I will return to this study below).

	Type of construction	Working class	Middle class	
			Informal	Formal
FRONTING	Où tu vas? Où Jean va?	36%	46	10
PRONOMINALISATION	Tu vas où? Jean va oú?	12	33	25
EST-CE QUE				
Est-ce que	Où est-ce que tu vas? Où est-ce que Jean va?	8	12	3
Que	Où qu' tu vas?	26	--	--
C'est que	Où c'est qu' tu vas	3	4	--
Que c'est que	Où que c'est que tu vas?	6	--	--
INVERSION				
Simple	Où vas-tu?		3	47
	Où sont passés mes cigares?	9	2	12
Copy of PRO	Où Jean va-t-il?	--	--	3
		(N-587)	N-446)	(N-436)

Table 1. Relative frequency of the distribution of interrogative constructions in a representative corpus of spoken French (Bernsted, 1973)

Two central elements of Labov's variationist model bear on issues in second language learning related to the Proficiency Movement. First, speakers' realizations of a variable are arranged along a continuum reflecting degree of attention paid to speech. Second, the vernacular pole, reflecting the least attention paid to speech and showing least variation, represents speakers' most regular and systematic linguistic patterns. The few available studies of variation in second language learners' interlinguistic production have adopted these two central hypotheses of Labovian sociolinguistics. L. Dickerson (1975), W. Dickerson (1975), and L. and W. Dickerson (1977) have examined the development of phonological interlinguistic features over time and in different situations of speech production. They show that in drawing closer to the TL model learners follow a course analogous to phonological change and to the synchronic variation found in NS speech: they produce a variety of replicas for given TL features, certain replicas predominating in a particular environment. Over time, and differentially across speech styles, replicas approximating the TL feature will spread progressively to new environments just as in a natural language standard variants spread differentially across various environments, word classes, social groups, or geographical areas.

On the basis of experimental studies focusing on morpho-syntactic as well as phonological features, E. Tarone (1979, 1983) hypothesizes that learners' interlanguage (IL) constitutes a continuum analogous to the Labovian vernacular-standard one. The relative position of an interlinguistic variant on the continuum is a function of the attention paid to linguistic form, see Figure 1. The left-most pole of the IL continuum corresponds to the vernacular on the natural language continuum. Vernacular interlinguistic features are produced when attention to linguistic form is lessened, and they exhibit greater regularity. On the contrary, the right-most pole, corresponding to the standard end of the native vernacular-standard continuum, is likely to contain TL and L1 features and shows the least internal consistency.

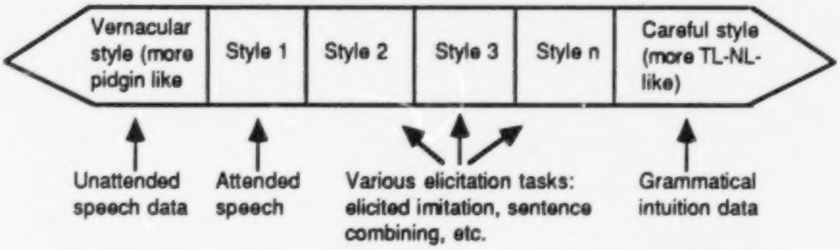

Vernacular style (more pidgin like
Style 1
Style 2
Style 3
Style n
Careful style (more TL-NL-like)
Unattended speech data
Attended speech
Various elicitation tasks: elicited imitation, sentence combining, etc.
Grammatical intuition data

Figure 1. The Interlanguage (IL) Continuum (from Tarone, E., 1983, p. 152)

Like Corder (1977) and Widdowson (1977), Tarone accounts for the internal consistency of the IL "vernacular" in terms of universal language processing strategies and substantive universals appearing under conditions of reduced input, such as those that trigger pidginization.

3. A multi-target model of language variation

3.1 Multi-target orientation in vernacular speech – Casual observation of speech use among speakers of the subordinate speech variety in a diglossic situation--monolingual speakers of Haitian Creole in Haiti, where the dominant language is French, for example--leads me to question the widely accepted early sociolinguistic model. According to that model all variation within a linguistic community may be located on a vernacular-standard continuum. In other words, there are no variants that reflect orientation toward some norm other than that which characterizes the speech of ENS. But in his study of urban Black English Labov (1972, p. 249) accounts for the persistence of stigmatized forms in the vernacular speech of Black youths by invoking a set of covert norms that attribute positive values to the vernacular. He also remarks that verbal behavior determined by these norms is unlikely to appear in the formal situations that characterize sociolinguistic interviews. The following case illustrates that non-standard speakers may be influenced by a variety of norms, including the prestigious standard form.

In casual observations of linguistic variation in a small village in the Touraine region of France, Pierre Léon (1973) noticed that a group of male speakers from farmer and worker backgrounds shifted away from local vernacular speech in two directions. For example, they replaced the local apical tap ɼ either by the velar resonant of standard French or by the pharyngal resonant characteristic of lower and working class Parisians. The latter feature coincided with macho behavior and took place typically at meetings in a café run by a retired railway worker from Paris, a former militant Communist union organizer. This social group's movement away from its vernacular proceeds in two different directions, one oriented toward ENS speech (the standard norm), and the other toward what may be termed Parisian proletarian speech. Shifts toward one or the other of these two vectors cannot be accounted for exclusively in terms of relative attention to linguistic form. The triggering factors are complex and must be sought in the nature of the social relationships among participants in the speech event and in the situational context.

As an alternative to the unidirectional vernacular-standard continuum model I propose a multi-target model, see Figure 2. I posit that all social groups in a speech community have available a range of speech styles and that these reflect the influence of several norms. In Figure 2 the square boxes labeled Norm A, Norm B, and Norm C, respectively, represent alternative norms available to the various social groups of a speech community. For the sake of simplicity the model posits only two social groups, X and Y. The width of the arrows reflect the relative power of attraction exercised on a particular social group by a given norm. Shifts in norm orientation are determined by the social identity speakers wish to signify in the course of given speech events. In daily communicative interactions with members of their own social groups, speakers make use of their vernacular, which is targeted on the norm set by members of the in-group. In complex societies one of the norms available in the overall speech community is endowed with prestige and/or political power. This norm undergoes a gradual progress of uniformization and gains the status of standard. It also becomes associated with the dominant social group whose vernacular will show less deviation from it than will that of the other social groups; in Figure 2, Group X represents the dominant group of the society whose speech networks are modelled. In the case cited by Pierre Léon (1973), the vernacular speech of local farmers is oriented toward the local rural Touraine norm, but they will shift toward the Parisian proletarian norm or Standard French, depending on the "acts of social identity" in which they engage in a particular situation (Le Page & Tabouret-Keller, 1984).

In addition to reflecting orientation toward a variety of norms a speaker's verbal production varies between planned and unplanned discourse. Unplanned discourse, which characterizes conversational interaction, shows the predominance of reliance on the

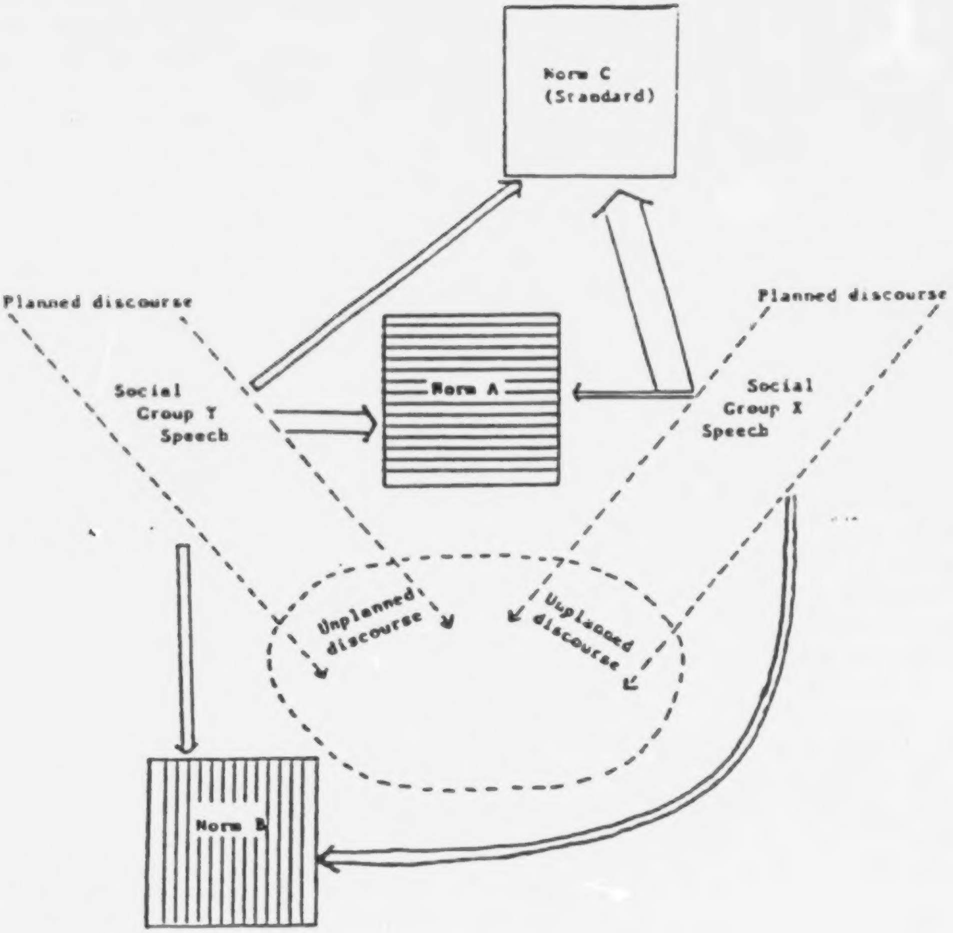

Norm C
(Standard)
Planned discourse
Planned discourse
Social
Group Y
Speech
Norm A
Social
Group X
Speech
Unplanned
discourse
Unplanned
discourse
Norm B

Figure 2. Multi-target model of language variation

situational context and on lexicon and, consequently, syntactic reduction (Ochs, 1979). Planning involves attention to speech form, so that the distinction I am making between planned and unplanned discourse corresponds somewhat to that made between attended and non-attended speech in Tarone's model (1983), see Figure 1 above. The chief difference between my model and hers lies in the independence as opposed to the convergence of two distinct continua: unplanned ➛ planned discourse (non-attentions ➛ attention to speech form) on one axis and vernacular ➛ external norm on another. By failing to recognize the independence of these two axes of speech variation linguists have assumed that non-standard speech lacks syntactic organization (Stourdzé, 1969; Bonin, 1978), and, as was the case in the Lindenfeld (1977) and Behnsted (1979) studies, they did not provide for the possibility that non-middle class speakers could produce both planned and unplanned discourse. As would be expected, because it relies on shared information and is linked to the hic et nunc of a specific communicative situation, the unplanned discourse of different social groups converge. Conversely, we would expect to find the greatest linguistic differentiation between two social groups to be located in their planned discourse.

3.2. Naturalness and the vernacular – Planned discourse and, in particular, standard norms involve a sort of complexification from a more regular and internally consistent "natural" linguistic base (Corder, 1977). This notion was first put forward by French linguists who initiated the serious study of so-called français populaire (lower and working class French; Bauche, 1929; Frei, 1929). Pierre Guiraud (1965) contrasted the naturalness and internal consistency of folk speech with the complexity and arbitrary nature of the planned discourse of ENS who focus on the standard norm:

> Le français populaire, en revanche, a conservé la simplicité, l'homogénéité, la vigueur et la naïveté d'une économie naturelle. ...le français cultivé est défini par des règles tirées à la fois d'une réflexion sur l'idiome et de l'expérience d'une tradition, alors que le peuple n'est soumis qu'aux lois naturelles qui gouvernent tout système de signes. "So called français populaire, on the other hand, has preserved the simplicity, the consistency, the vigor and the pristine quality of a natural economy. ...learned French is defined by rules derived at the same time from reflection on the language and from the experience of a tradition, while the folk is governed only by the natural laws which characterize all systems of signs" (1965, p. 12).

3.3. A multi-target norm for IL variation – Within the framework of the multi-target norm model I propose for TL variation, the IL continuum may be viewed as a vector which is oriented toward a particular TL norm by filtering input and controlling feedback, see Figure 3. The largest square represents the totality of attested TL variants.

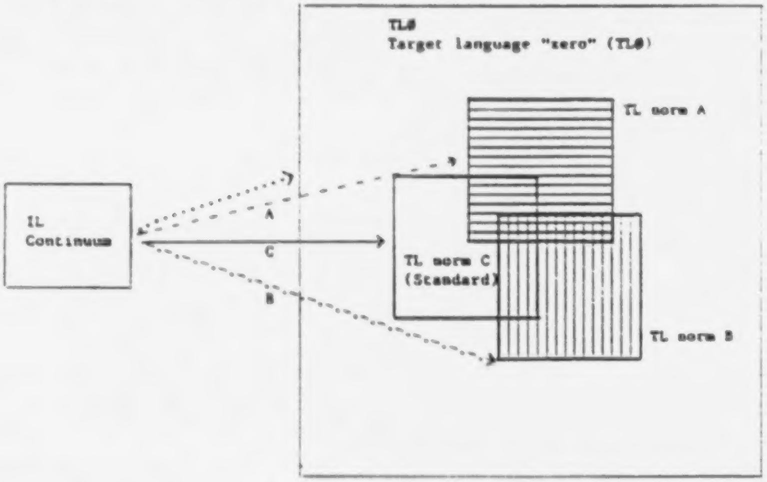


Figure 3. Possible orientations of IL continuum toward various TL norms

Following R. Chaudenson (1986) I call this system TL_0. This system encompasses features shared by all TL varieties as well as all variable features. The three smaller squares represent competing norms, one of which is the standard. In naturalistic second language learning the types of communicative situations encountered by the learner will determine in large part the orientation of the IL continuum. In guided learning it may be controlled to a certain extent by the selection and sequencing of linguistic features, by classroom procedures, and more particularly by the elaboration of pedagogical norms.

4. The elaboration of pedagogical norms

4.1. The concept of pedagogical norm – There is a growing recognition on the part of second language acquisition researchers that the foreign classroom constitutes a special environment for second language development (Ellis, 1984; Faerch & Kasper, 1985; Edmonson, 1985; Kramsch, 1985). The varieties of the TL that are appropriate for classroom instructional and communicative activities are determined by the attitudes of the TL community, by those of the learners themselves, as well as by factors inherent in these activities (Soulé-Susbielles, 1984).

Whereas there exists a considerable body of research on native speakers' reactions toward learners' interlinguistic features (see d'Anglejan, 1983 for a review of the literature), socio-psychologically oriented studies of native speaker attitudes have not yet been extended to the area of language variation and the definition of a sociolinguistically acceptable learner repertoire for the TL. From the only available exploratory study (Swacker, 1976), it would appear that native speakers tend to downgrade foreigner speech that diverges markedly from the standard norm. Swacker observed that native speakers of East Texas English reacted negatively to a reading sample from an Arabic accented speaker whose speech contained local pronunciation and grammatical features. A higher ranking was assigned to another Arabic accented speaker whose English was free of local linguistic features. She concluded that certain regional markers may be acceptable when found in the speech of native speakers but are offensive in the mouth of foreigners. Swacker fails to specify, however, whether the local features noted also symbolize low social status within the TL community.

Without a substantial body of empirical research we must resort to extrapolation from other areas of interlinguistic research to seek principles for the determination of TL norms suitable for foreign learners. The reaction of Swacker's native judges may be accounted for in terms of a model of status and solidarity (Ryan, 1983). Foreign accents and other interlinguistic features are highly salient markers of membership in outgroups. Their presence in the speech of a foreigner would preclude features symbolizing ingroup membership, such as regionally or socially marked variants. Appeal to the notion of linguistic capital (Bourdieu, 1977) also supports the selection of a highly standardized TL norm for foreign learners. The learning of the TL may be viewed as an economic investment whose value would be depreciated if the variety mastered contained stigmatized features. The higher the social status conferred by the learned speech forms and the communicative strategies, the more profitable the investment.

These various considerations suggest that the language used in the classroom should differ from the naturalistic varieties found in communicative interactions among native speakers. The differences between what may be termed a pedagogical norm and the various norms that determine which native speaker usage will be particularly marked at early stages of instruction when learners' speech shows many deviant features, some attributable to transfer from the native tongue, others reflecting extensive restructuring of TL input, and still others residing in the communicative strategies specific to the classroom context. Thus, a central task in organizing classroom FL instruction is the elaboration of pedagogical norms, particularly at early stages.

A pedagogical norm, like all norms, is an abstraction. Its distinguishing feature resides in the fact that it is an artificial construct reflecting the special conditions of classroom FL learning (Valdman, 1976, 1983). Four sets of principles guide the elaboration of pedagogical norms. First, they should reflect the actual behavior of TL speakers in authentic communicative situations. Second, they should conform to native speakers' idealized views of their linguistic behavior. Third, they should match both TL speakers' views about what linguistic behavior is appropriate for educated foreigners and the perceptions of learners themselves. Fourth, they should take into account processing and learning factors, such as learners' interlanguage features, since these point the way to what is most easily learnable.

The elaboration of a suitable pedagogical norm requires first of all a clear determination of learning objectives and a careful review of the social and broad contextual factors that are likely to shape the particular classroom environment. From a more narrowly linguistic and psycholinguistic perspective it requires linguistic descriptions of a level of delicacy not attainable by the application of current linguistic models. Not only must applied linguists be supplied with information on the semantic, pragmatic, and sociolinguistic value of specific target language features but they must also be given insight into the relative optimality of the available structural options. I am using the term optimality to refer to ease of acquisition and centrality within the overall structure of the TL. This notion no doubt corresponds in part with markedness in the sense that optimal features would be expected to be unmarked from the point of view of language universals.

4.2 Illustration of a pedagogical norm – Limitations of time do not permit a detailed illustration of the notion of pedagogical norm. I will only summarize its application to a notoriously variable aspect of French syntax--WH-questions--I have treated elsewhere (Valdman, 1975, 1976, 1983, 1986).

Traditionally, two WH-constructions are selected for the teaching of French as a foreign language: EST-CE QUE and INVERSION. However, as Behnsted's sociolinguistically oriented study shows (see Table 1), these constructions are relatively rare in vernacular speech. Instead native speakers use two variants often characterized as "ungrammatical", but which are syntactically less complexified, PRONOMINALIZATION and FRONTING. It is noteworthy that ENS (Behnsted's middle class informants) overestimated their use of the most valorized construction, INVERSION, and, inversely, underestimated the frequency of the two devalorized variants.

There are no studies of NS attitudes regarding the relative appropriateness of the WH-question variants for foreign learners. Extrapolating from Swacker's pilot study (1976) and referring to the criteria of status and solidarity (Ryan, 1983), one may assume that French ENS would expect educated native speakers to adhere closely to their own subjective norm and favor the two constructions associated with formal styles. From a psycholinguistic perspective, data from two different lines of inquiry suggest that the two most frequent WH-question types--PRONOMINALIZATION and FRONTING--are the most easily learnable. Firstly, the latter construction is universally found in European-based creoles where vernacular varieties of the lexifier language show different syntactic arrangements than the creole. For example, compare Where have you been? and Guyanese Creole Wisaid yu bin de? (Bickerton, 1981, p. 70). Secondly, a study of the acquisition of French WH-question by American English learners (Valdman, 1975, 1976) revealed a high proportion of FRONTING. That structure occurred in utterances produced within the framework of a test of oral ability administered to beginners after approximately 75 hours of instruction. The presence of the structure represents a case of creative construction (Burt, Dulay & Krashen, 1982) since the learners had never been exposed to it; the construction was absent from the course materials and from the speech of instructors, all of whom used INVERSION or EST-CE QUE, in both drill and communicative activities. The presence of the feature can not be explained by transfer from the L1 either since English WH-questions may be described, at the surface level, either by inversion or by the adjunction of a semantically empty marker do matching EST-CE QUE with respect to position:

Where	does	John	work?
↕	↕	↕	↕
Où	est-ce que	Jean	travaille?

The application of the four criteria proposed for the elaboration of pedagogical norms suggests the following dynamic norm for the introduction of WH-questions. The norm is dynamic in the sense that the learner's linguistic behavior would be expected to progressively draw closer to the selected target. The first three criteria for a pedagogical norm converge to identify INVERSION as the most highly valued structure from a sociolinguistic point of view. No doubt native speakers would deem it most appropriate for foreigners who have acquired the language by formal instruction. Bourdieu's view of language as capital would also lead learners to assign it the highest surrender or market value. On the other hand, it is highly marked, both from the syntactic and the sociolinguistic point of view. Its close link with formal discourse and written texts render it inappropriate as the target feature in elementary and intermediate courses which have as one of their main objectives the acquisition of a certain degree of communicative ability. The numerous syntactic constraints that apply to INVERSION (Valdman, 1975) render it difficult to learn and to use with accuracy. Its sociolinguistically neutral status and its relative syntactic simplicity designate EST-CE QUE as the first-level target feature. However, since FRONTING represents a case of creative construction it is suitable for transitional use in early stages of instruction, see Figure 4. INVERSION would be introduced at Stage 2 in lexicalized frozen forms in cases where EST-CE QUE is judged too awkward, for example compare: Comment allez-vous? 'How are you?' (Comment est-ce que vous allez?), Où vont vos parents? 'Where are your parents going?' (Où est-ce que vos parents vont?). At that stage INVERSION would also serve as the canonical form for written expression. At more advanced levels (Stage 3) PRONOMINALIZATION and FRONTING would be identified as more appropriate for informal situations. As is the case for INVERSION, their use in contrived materials should reflect their distribution in authentic, unedited speech samples so that learners may be led to induce their pragmatic and sociolinguistic values. Figure 4 represents a proposed pedagogical progression indicating the order of introduction of the four variants in speech and writing in which a distinction is made between receptive and productive phases.

As was pointed out in Section 3.3 above the concept of pedagogical norm makes it

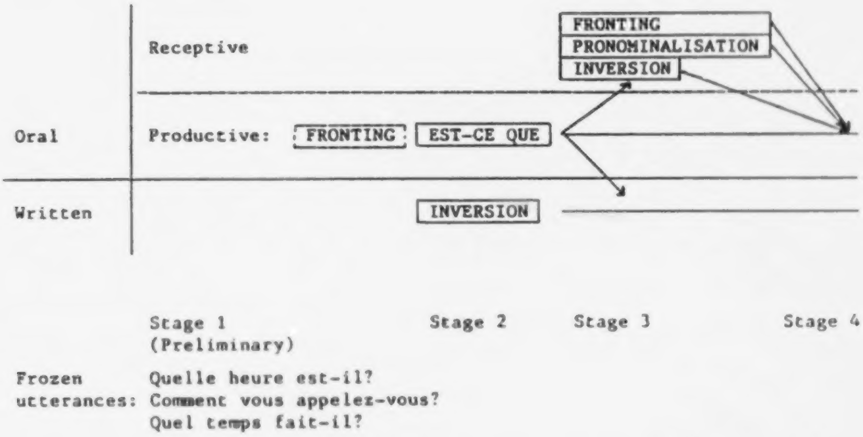


Figure 4. A proposed pedagogical sequence for the presentation of French interrogative constructions

possible to progressively orient the IL vector toward the ultimate TL norm selected. The TL variant selected for initial presentation is the one that learners can most easily process. Should it be a stigmatized feature, it is progressively replaced by a sociolinguistically neutral or valorized variant. In the specific case of French WH-questions, FRONTING would be selected as the first target feature. It would be replaced by EST-CE QUE as soon as the learners would have acquired control of WH-words. INVERSION would be introduced in written discourse and, in the final stage, all four constructions would be presented for recognition and active control in appropriate situational contexts. Any particular group of learners would need to demonstrate the ability to fit the WH-question variant construction to the intended audience and the specific situational context. Suppose that the instructional objective were to impart a level of linguistic proficiency sufficient to enable Foreign Service officers to perform Level 4 tasks, these learners would need to attain Stage 4 in Figure 4. On the other hand, the IL vector would need to be set differently in the case of the average high school or college level student for whom Stage 2 or early Stage 3 would be sufficient to carry out Level 2 tasks, for example, to engage in all types of casual conversations.

5. The issue of fossilization

A major issue in the Proficiency Movement is that of fossilization. Higgs and Clifford (1982) report that many trainees in governmental agencies' language programs who begin instruction with a relatively high level of proficiency (S-1+) fall short of attaining the critical S-3 level because lexical and grammatical structures imperfectly mastered during previous study or language use become fossilized. To remedy this situation, they propose more stringent attention to linguistic accuracy at initial levels of second language development. From that perspective, one could object to the application of pedagogical norms in which learner performance is oriented toward features characteristic of unplanned and non-standard vernacular models because it could lead to the fossilization of these features. As a result, learners would fail to attain the control of variants characteristic of ENS whose planned speech is generally targeted on the standard norm. This type of fossilization would constitute a serious problem in the training of diplomats whose job description required negotiating with ENS counterparts.

I would respond to this type of objection by underscoring the distinction between IL features that fall outside of the construct I have labeled TL_0 (see Figure 3) and those within it. The former features are indeed linguistic inaccuracies to which the standard notion of fossilization applies. Sociolinguistically stigmatized features, which abound in the speech of NS, may at worst be sociolinguistically or discursively inappropriate. And sociolinguistic and discursive inappropriateness also result when variants that characterize planned, formal discourse occur in situations that require unplanned discourse; such is the case, for example, for French INVERSION WH-questions. It is only the traditional enshrining of ENS planned discourse which leads to labeling features characteristic of non-standard speech and unplanned discourse (often termed slip-shod) as grammatically inaccurate. In addition, discussions of fossilization in IL fail to distinguish genuine instances of this phenomenon from the syntactic reduction linked to reliance on the situational context that accompanies shifting down from planned to unplanned discourse. In the context of the Oral Proficiency Interview, one wonders whether some cases of fossilization are not an artifact of the testing situation. At lower proficiency levels the OPI can at best simulate naturalistic conversational interaction. Genuine conversational interaction is more likely to elicit unplanned discourse. Indeed, planned discourse would be artificial in that type of situation.

If anything, reliance on the situational context and on lexicon and accompanying reduction of syntax are pragmatically compatible and sociolinguistically appropriate in a test portrayed as an approximation of genuine conversational interaction. It may be that learners afflicted with the so-called terminal 2/2+ syndrome might perform with greater grammatical accuracy in test situations that more closely model situations requiring the production of planned discourse.

6. Conclusion

Higgs and Clifford (1982, p. 78) propose the "output hypothesis", analogous to Krashen's optimal input hypothesis (1982):

> When the communicative demands made upon the students are too far beyond their current competence, they are forced to adopt or invent communication strategies that lead to fossilization and ultimately prove self-defeating. For this reason it is important to match the communicative task to the students' performance level.

I would agree that such mismatches between communicative and semantic intent and linguistic form must be avoided. However, the converse to the "output hypothesis" is equally dysfunctional:

> When second language learners are forced to use grammatical features and to make semantic differentiations that are beyond their processing capabilities, they resort to structural reduction. This leads to the production of features, characteristic of

> unplanned discourse that appear structurally deviant
> from the perspective of ENS planned discourse.

It is the responsibility of syllabus designers, developers of teaching materials, and classroom instructors alike to avoid both types of mismatches between language use tasks and linguistic means available to learners. The notion of pedagogical norm I have developed in this paper promises to lead to this mutual adjustment between language form and language use.

Schulz (1986) voices serious reservations about the appropriateness of oral proficiency in a foreign language as a central goal in general education or liberal arts training. She argues that, in addition to the fact that it can only be attained in extensive sequences of study not usually found in U.S. schools and universities' speaking ability is the language skill most vulnerable to erosion or loss. On the basis of these and other considerations I proposed (Valdman, 1982) that in addition to broad educational objectives, such as acquainting students with the multifarious links between language and mind, culture, and society, fundamental (as versus special purpose) foreign language teaching instruction should more modestly and realistically aim at imparting minimal communicative capacity, MCC, rather than full-fledged communicative proficiency.

Within the framework of this type of fundamental foreign language instruction the elaboration of pedagogical norms is all the more crucial. Clearly, MCC cannot be attained in a classroom context by the average learner if ENS planned discourse remains the target. That variety of the target language shows the highest level of syntactic and semantic complexification. At the level of syntax many minor rules concerning subcategorization, government, and agreement must be acquired; and since speech will be often decontextualized, highly specific vocabulary items expressing fine nuances of meaning must be introduced. In contrast unplanned discourse, which is highly context-embedded, features syntactic reduction and semantic simplification. A pedagogical norm comprises target language variants that involve the simplest syntactic machinery and that form the most regular pattern. Where social or geographical diversity exists sociolinguistic factors will be abandoned in favor of psycholinguistic ones: those variants will be selected that are most easily processable by second language learners. It should be stressed that this approach does not involve the teaching of a pidginized version of the TL. The features that constitute the pedagogical norm do not have their origin in learners' interlinguistic approximations; they are drawn from variants occurring in a composite of actual native speech.

The approach I am advocating takes a neutral stand with respect to current debates about suitable classroom activities -- communicative practice or modelling of a set of selected structural features -- and about appropriate syllabus design -- notional-functional or structural. For example, most speech acts may be expressed either by linguistic features characteristic of ENS or those of other less prestigious social groups; the same set of semantic notions may be overtly realized by the complex syntactic machinery typical of planned discourse or by the sentence fragments, formulaic combinations, and unmarked syntactic devices and morphological categories characteristic of unplanned discourse. The bottom line is that learners should be led to attain MCC as quickly as possible but in a manner that permits, in the long term, the development of language-specific marked linguistic devices that make decontextualized discourse possible. This objective will not be reached without a drastic overhaul of the linear structurally oriented syllabus that underlies most current teaching, including those that claim to be "proficiency-oriented".

Note

[1] Lowe (1986) points out that in the parent ILR (Interagency Language Roundtable) the upper end of the scale is labeled well-educated native speaker (WENS). There does not seem to be any significant difference between the two concepts.

References

Behnsted, P. (1973). Viens-tu, est-ce que tu viens? *Formen und Strukturen des direkten Fragesatzes im Französischen.* Tübingen: Gunter Narr.

Bengtsson, S. (1968). *La défense organisée de la langue française.* Upsala.

Bickerton, D. (1981). *Roots of language.* Ann Arbor, MI: Karoma Press.

Bonin, T. (1978). The role of colloquial French in communication and implications for language instruction. *Modern Language Journal, 62,* 90-102.

Bourdieu, P. (1977). L'économie des changements linguistiques. *Langue française, 34,* 17-34.

Chaudenson, R. (1986). Norme, variation, créolisation. *AILA Review/Revue de l'AILA, 2,* 69-88.

Chaudenson, R., Valli, A., & Véronique, D. (1986). The dynamics of linguistic systems and the acquisition of French as a second language. *Studies in Second Langage Acquisition, 8,* 277-292.

Corder, S. P. (1976). 'Simple codes' and the source of the second language learner's initial heuristic hypothesis. *Studies in Second Language Acquisition, 1,* 1-10.

Corder, S. P. (1977). Language continua and the Interlanguage Hypothesis. In S. P. Corder & E. Roulet (Eds.), *Actes du 5ème Colloque de Linguistique Appliquée de Neuchâtel: The notions of simplification, interlanguages and pidgins and their relation to second language pedagogy* (pp. 11-17). Geneva: Librairie Droz.

D'Anglejan, A. (1983). Introduction to special issue, native speaker reactions to learner speech. *Studies in Second Language Acquisition*, 5(2), vii-ix.

Douglas, D. (1986). *Communicative competence and tests of oral skills.* In C. W. Stansfield (Ed.), *Toward communicative competence testing.* Proceeding of the Second TOEFL Invitational Conference (pp. 156-174). TOEFL Research Report 21. Princeton: Educational Testing Service.

Dulay H., Burt, M., & Krashen, S. (1982). *Language two.* New York: Oxford University Press.

Edmonson, W. J. (1985). Discourse worlds in the classroom and foreign language learning. *Studies in Second Language Acquisition, 7,* 159-68.

Ellis, R. (1984). *Classroom second language development.* Oxford: Pergamon Press.

Faerch, C., & Kasper, G. (1985). Introduction to special issue, Foreign language learning under classroom conditions. *Studies in Second Language Acquisition, 7,* 131-133.

Fishman, J. A. (1967). Bilingualism with and without diglossia; diglossia with and without bilingualism. *Journal of Social Issues, 23,* 29-38.

Frei, H. (1929). *La grammaire des fautes.* Paris: Geuthner.

Guiraud, P. *Le français populaire.* (Que Sais-je No. 1172). Paris: Presses Universitaires de France.

Heilenman, L.K., & Kaplan, I. (1985). *Proficiency in practice: The foreign language curriculum.* In C. J. James (Ed.), 55-78.

Higgs, T. V., & Clifford, R. (1982). The push toward communication. In Higgs, T. V. (Ed.), *Curriculum competence and the foreign language teacher* (pp. 57-80). Skokie, IL: National Textbook Company.

James, C. J. (Ed.). (1985). Foreign language proficiency in the classroom and beyond. Lincolnwood, IL: National Textbook Company.

James, C. (1985). Learning from proficiency: The unifying principle. In C. J. James (Ed.), 1-8.

Kramsch C. J. (1985). Classroom interaction and discourse options. *Studies in Second Language Acquisition, 7,* 169-183.

Krashen, S. (1982). *Principles and practice in second language acquisition.* Oxford: Pergamon Press.

Labov, W. (1966). The social stratification of English in New York City. Washington, DC: Center for Applied Linguistics.

Labov, W. (1973). *Sociolinguistic patterns.* Philadelphia: University of Pennsylvania Press.

Léon, P. R. (1973). Réflections idiomatologique sur l'accent en tant que métaphore sociolinguistique. *French Review, 46,* 783-789.

Lowe, P. Jr. (1985). The ILR proficiency scale as a synthesizing research principle: The view from the mountain. In C. J. James (Ed.), 9-54.

Lowe, P. Jr. (1986). Proficiency: Panacea, framework, process? A reply to Kramsch, Schulz, and, particularly, to Bachman and Savignon. *Modern Language Journal, 70,* 392-397.

Ochs, E. (1979). Planned and unplanned discourse. In T. Givòn (Ed.), *Syntax and Semantics: Vol. 12. Discourse and semantics.* New York: Academic Press.

Ryan, E. B. (1983). Social psychological mechanisms underlying native speaker evaluations of non-native speech. *Studies in Second Language Acquisition, 5,* 148-159.

Savignon, S. J. (1986). The meaning of communicative competence in relation to the TOEFL program. In C. W. Stansfield (Ed.), *Toward communicative competence testing* (TOEFL Research Report no. 21, pp. 17-30). Princeton, NJ: Educational Testing Service.

Schultz, R.A. (1986). From achievement to proficiency through classroom instruction: Some caveats. *Modern Language Journal, 70,* 373-379.

Soulé-Susbielles, N. (1984). La question, un outil pédagogique dépassé? *Le français dans le monde, 183,* 26-34.

Stansfield, C. W. (Ed.). (1986). Toward communicative competence testing. Proceedings of the Second TOEFL Invitational Conference (Research Reports 21). Princeton, NJ: Educational Testing Service.

Stourdzé, C. (1969). Les niveaux de langue. *Le français dans le monde, 65,* 18-21.

Swacker, M. (1976). When (+native) is (-favourable). *Lektos* [special issue], 135-254.

Tarone, E. (1979). Interlanguage and the chameleon. *Language Learning,* 29, 181-191.

Tarone, E. (1983). On the variability of interlanguage systems. *Applied Linguistics, 4,* 149-163.

Trevise, A., & Porquier, R. (1986). Second-language acquisition by adult immigrants: Exemplified methodology. *Studies in Second Language Acquisition, 8,* 65-76.

Valdman, A. (1972). Language variation and the teaching of French. In G. Todd (Ed.), *Current issues in the teaching of French* (pp. 87-108). Philadelphia: Center for Curriculum Development.

Valdman, A. (1973). La contribution de la linguistique appliquée américaine à l'élaboration de progressions motivées. *Etudes de Linguistique Appliquée, 10,* 53-73.

Valdman, A. (1975). Error analysis and pedagogical ordering. In S. P. Corder & E. Roulet (Eds.), *Some implications of linguistic theory for applied linguistics* (pp. 105-26). Paris: Didier.

Valdman, A. (1976). Variation linguistique et norme pédagogique dans l'enseignement du français langue étrangère. *Bulletin de la Féderation Internationale des Professeurs de Français, 12-13*, 52-64.

Valdman, A. (1980). Communicative ability and syllabus design for global foreign language courses. *Studies in Second Language Acquisition, 3*, 81-96.

Valdman, A. (1982). Toward a modified structural syllabus. *Studies in Second Language Acquisition, 5*, 34-51.

Valdman, A. (1983). Language variation and foreign language teaching: Issues and orientations. In L. MacMathuna & D. Singleton (Eds.), *Language across cultures*, (pp. 171-84). Proceedings of a symposium held at St. Patrick's College, Drumcondra, Dublin: Irish Association for Applied Linguistics.

Widdowson, H. G. (1976). The significance of simplification. *Studies in Second Language Acquisition, 1*, 11-20.

DISCUSSION:
FOUR CHALLENGES FOR PROFICIENCY: COMMENTS FROM WITHIN THE AEI PROFICIENCY FRAMEWORK

Pardee Lowe, Jr.
Language Training Division
Office of Training & Education
The Central Intelligence Agency

1. Introduction

Prof. Valdman presents an admirably balanced view of the issues surrounding the interlanguage continuum, target language standards--including the Educated Native Speaker (ENS) and other norms--as well as their relation to accuracy and fluency in assessing and developing target language proficiency. Prof. Valdman's paper, however, also raises four larger issues, which I wish to address from within the AEI proficiency framework. The designation AEI subsumes those features shared by the ACTFL/ETS and ILR scales. For example, the threshold nature of the scales' borders and the gestalt nature of rating student performances. Where the ACTFL/ETS and ILR systems diverge, the AEI designation is not applied (Amplified in Lowe, 1986, pp. 391-397).

2. The four challenges

Four challenges await us: defining terms, encompassing variety, specifying the AEI scales' developmental hierarchy more fully, and deriving pedagogical norms.

The first challenge: Defining terms – As the French say, "Ça va sans dire." Yet, we often fail to define terms, purposely leave them vague, or fashion the broadest possible definition as Larson and Jones (1984) suggest. The consequence of terms left un- or ill-defined, is that we may not know to what we really refer when comparing "proficiency-oriented" assessment, curriculum design, and research to other approaches. Or worse, we may simultaneously contend with several varieties of proficiency, insufficiently identified or differentiated, such as AEI proficiency and Cummins' concept (1985) which is based in part on a developmental hierarchy for non-adults (see below) or Bachman/Savignon's (1986) which weights language for special purposes more highly than does AEI proficiency. To differentiate AEI proficiency from other varieties, we hazard the following operational definition:

> AEI proficiency equals achievement (ILR functions, content, accuracy) plus functional evidence of internalized strategies for creativity expressed in a single global rating of general language ability over a wide range of ILR functions and topics and any given ILR Level (Lowe, 1985b, p. 16f)

Other terms used, such as fluency and accuracy, should also be defined. Of course, we might each define them differently, and then agree to disagree. For fluency--one might employ the interagency definition:

> The *fluency-integrative* factor combines both speech rate when compared to an ENS and the ease with which the elements of speech are put together (Lowe, 1985a, pp. 4-8).

And for accuracy:

> This factor refers to acceptability, quality, and accuracy of the message conveyed (Lowe, 1985a, pp. 2-3, 7-8).

In the course of the symposium we have had occasion to deal with accuracy from its narrowest possible definition to its broadest. Since Prof. Valdman uses the term (and I have employed it myself elsewhere), I refer to it here. But some *larger concept* is clearly in order, as the following story suggests.

Some time ago a man who owed me an achievement test stopped by to apologize for the test's being late. As a non-native speaker of English at the upper-end of the lower AEI spectrum, he attempted to shift registers to convince me of his embarrassment:

> "I am exceedingly occupied. I must relieve myself. I cannot disperse myself everywhere!"

To this day, he remains unwitting of what he said, and the situation is such that one cannot correct his error. His morphology and syntax are flawless. And, of course, we can reconstruct the intended message:

> "I'm up to my ears in work. I have to cut down somewhere. I can't have a finger in every pie."

Of interest here is the inappropriacy of the message conveyed. He put his idea across--and something more! Thus to encompass such speech, we require a concept larger than grammatical accuracy, or accuracy in general. To this end, Byrnes (1986, p. 110) employs the designation "appropriate," and I commend that usage to our consideration. Lesser concepts must also be defined. The drafters of the ACTFL/ETS Guidelines encountered problems with terminology when working with the writers of the ILR Skill Level Descriptions: in ILR usage, for example, "global" understanding comprises comprehending 70-80% of a passage; in academe "global" only comprises comprehending the main ideas. These examples suggest that to describe AEI proficiency more accurately and to communicate more readily with one another-- we must turn our hand to definition.

The second challenge: Encompassing variety – There awaits a second, greater challenge: specifying more fully how the AEI framework encompasses variety. Most of the contributions to the symposium, it seems to me, pose this challenge in one form or the other. Moreover, there exists a clear need for the profession as a whole to wrestle with the problem.

LEVEL

5

CAREFUL UNPLANNED SPEECH CAREFUL PLANNED SPEECH

4

UNPLANNED PLANNED

3

UNPLANNED PLANNED

= =

2

TERMINAL

STREET CLASSIC SCHOOL

1

TERMINAL

STREET CLASSIC SCHOOL

0 +

STREET

SOME VARIEITIES OF SPEECH ON THE ILR SCALE

FIGURE 1

Yet, this challenge is undertaken at some risk, for we must specify more fully the variety the AEI framework handles in such a way as to simultaneously preserve AEI proficiency for its original purpose, assessment.

Another facet to the challenge is discovering a way in which experienced testers can share with others the extent to which AEI proficiency already encompasses a wide variety of speech behaviors.

Figure 1 displays some varieties of speech already identified through AEI proficiency testing. The terms may require definition. From right to left, "school" learners are formal learners whose grammar is solid and whose everyday vocabulary is limited. A "classic" speaker evinces a balance between solid grammar and useful vocabulary. "Street" learners are those who learned the language informally and whose everyday vocabulary--due to exposure in country--is existent, weak, or fossilized. A "terminal" speaker is an exaggerated variety of "street," one whose language is unlikely to be remediated. Except for the "classic" speaker, these designations in ILR usage, cannot apply above the Level 2+. The levels above 2+ are non-compensatory and, thus, any global rating must by based on a common floor in these factors (For details, see Higgs & Clifford, 1982, p. 65ff; Lowe, 1985b, pp. 3-19ff).

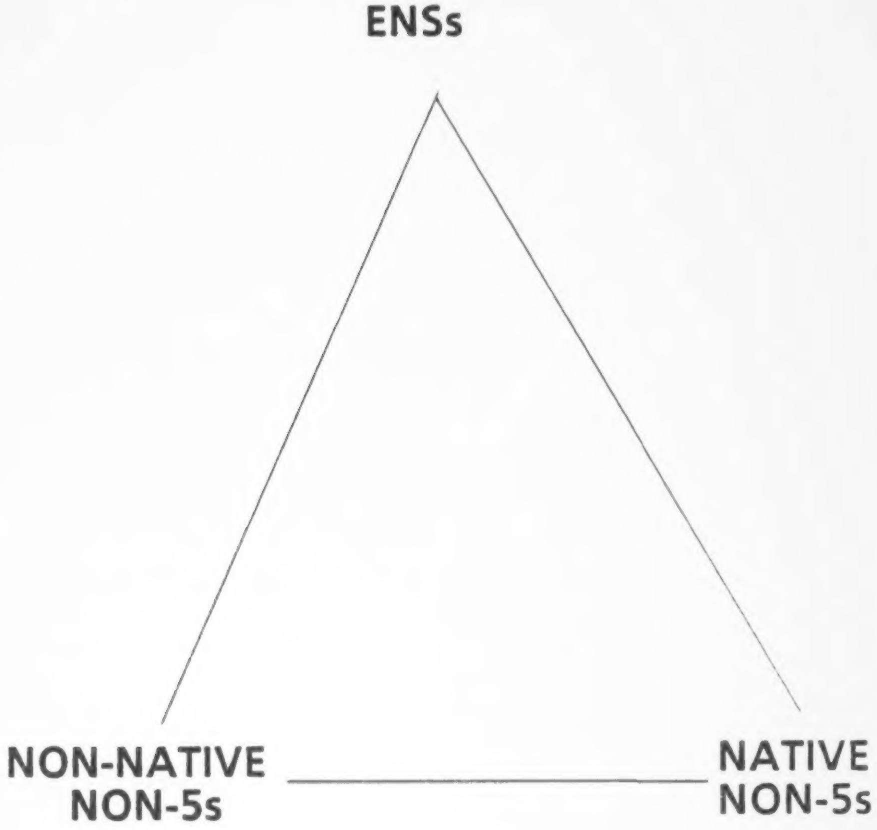
At Level 3 and higher I have used Prof. Valdman's terms, "unplanned" and "planned." If I understand them aright, they may, unlike the ILR terms, apply throughout the scale. Juxtaposed with the ILR terms as they are in Figure 1, they provide a fuller description of the variety that experienced testers report. There may appear to be a contradiction at Level 5 between the qualifiers "careful" and "unplanned." Applying both to one type of speech attempts to capture that "unplanned" speech one hears on occasions such as press briefings where questions are sometimes answered off the cuff. Despite the spur-of-the-momentness of the answer, it is evident as one listens that heavy monitoring contributes to the final product--hence, "careful unplanned speech."

How may we capture the variety already encompassed by the system? To this end, we could prepare protocols to reveal the kinds of target language candidates use in oral interviews (or write, or can read, or can listen to and comprehend). Subsequently, interviews with raters would examine the reasons such samples were assigned particular ratings.

It is within the context of the wide variety the AEI framework already encompasses, that I see the ENS (Educated Native Speaker). Indeed, the concept is--as Prof. Valdman indicates--an oversimplification. On the other hand, for some types of work, such as representing the US government abroad, "the slightly formal foreigner" may prove preferable to "the Ugly American." At some point, we must raise our sights from delight in any communication whatever to a consideration of the effects of communicating less or more than originally intended, as in the case of the apology related above. In considering the ENS' position within the AEI scales, it is important to stress that the AEI framework was designed to assess the communicative ability of Americans who had learned foreign languages to varying degrees of proficiency. The ENS tops the scale as a reference point. That the ENS has a trickle-down effect, no one would deny. But the effect may prove less than some have supposed (Savignon, 1985, p. 131f). What supplements the ENS concept? Experienced testers cite three groups of samples (See Figure 2):

1. "Classic" samples of ENS performances at the top of the scale; 2. "Classic" samples of native non-5 speakers elsewhere on the scale; 3. "Classic" samples of non-native non-5 speakers elsewhere on the scale.

Besides comparing performances to those described in the definitions, experienced


ENSs
NON-NATIVE NON-5s
NATIVE NON-5s

TRIANGULATION

FIGURE 2

raters report that they "triangulate" new performances against these three sample groups. "Classic" samples prove hard to find. But once encountered--testers refer to them constantly. We turn now to the third challenge, the nature of the AEI developmental hierarchy and its relation to acquisition research.

The third challenge: Specifying the AEI developmental hierarchy more fully – There exists a natural tension between empirically based and experientially based research. This is nowhere more evident than in discussing the AEI scales and their application. The ILR's 30-year long experience suggests that the ILR Skill Level Descriptions contain a developmental hierarchy. How such a hierarchy is to be described and objectified presents the third challenge.

The AEI Scales and adults: In the AEI scales "developmental" refers to stages of development in language use with those learning and using a language in their adult years. In other words, what occurs when an adult learns a brand-new language? Do there exist commonly recognized stages, generally applicable across numerous languages? In this sense, the AEI scales describe, generally, easier tasks before harder ones. In a way, one can understand this hierarchy much as one understands language universals: a pattern revealed in many, but not all, languages. This means that individual languages may prove exceptions, but most will not. It also means that the AEI developmental hierarchy reflects some larger abstraction which differs both from the developmental hierarchy of any individual language, and from the learning acquisition sequence characteristic of that language. What relationships obtain between the AEI developmental hierarchy in the widest sense and a specific language's acquisition sequence in the narrowest sense? This constitutes virgin territory.

The AEI Scales and non-adults: The ILR scale was devised to describe the proficiency of adult foreign language users. The ACTFL/ETS extension of the scale is also being used with foreign language users who are younger. What remains constant, what differs in such applications? One clear implication is that the AEI scales are integrative; that is, at the mid and higher levels, the scales require the integration of a number of skills and areas beyond those regarded as purely linguistic: culture, real-world knowledge, the ability to think, to name a few. What happens when the scales are applied to those who are younger and may lack fully developed ancillary skills?

The fourth challenge: Deriving pedagogical norms – All these considerations culminate in the fourth challenge, Prof. Valdman's call for deriving pedagogical norms. As linguists desirous of accurately describing language acquisition and as instructors desirous of producing proficiency at functional levels, we are naturally interested in the curricular import of different paths to the same goal. With limited teaching time we seek the most efficient paths.

Prof. Valdman's suggestions in this regard merit serious study and deserve fuller application. Among the numerous questions he raises, I would pursue these: Must we always proceed through TL_O? What can be skirted in TL_1? These ideas should be applied to a larger number of structures in French and be extended to structures in other languages as well.

Perhaps, beyond the numerous questions Valdman raises, beyond the colloquial norm, beyond sociolinguistic variety, beyond the ENS, lurks a more general question: What variety(-ies) offer(s) the greatest appropriacy in the teaching time available? Again, such a question can be answered empirically. Because the AEI scales rate growing "independence," the question might be rephrased, "What variety (-ies) and approaches most promote such independent language use?"

3. Conclusion

I began by citing the four challenges for proficiency which Prof. Valdman's paper raises. There would be no more fitting conclusion than to tackle them!

References

American Council on the Teaching of Foreign Languages. (1986). *ACTFL guidelines: Generic descriptions*. Hastings-on-Hudson, NY: Author.

Bachman, L. F., & Savignon, S. V. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. *Modern Language Journal, 70*, 391-97.

Byrnes, H. (1986). Second language acquisition: Insights from a proficiency orientation. In M. Canale and H. Byrnes (Eds.), *Defining and developing proficiency: Guidelines, implementations, and concepts* (pp. 107-31). Lincolnwood, IL: National Textbook.

Cummins, J. (1984). Wanted: A theoretical framework for relating language proficiency to academic achievement among bilingual students. In C. Rivera (Ed.), *Language proficiency and academic achievement* (pp. 2-19). Clevedon/Avon, England: Multilingual Matters.

Higgs, T. V., & Clifford, R.T. (1982). The push towards communication. In T. V. Higgs (Ed.), *Curriculum, competence, and the foreign language teacher* (pp. 57-79). Lincolnwood, IL: National Textbook.

Government Interagency Language Roundtable. (1985). *ILR skill level descriptions*. Washington, DC: Author.

Larson, J. W., & Jones, L. (1984). Proficiency testing for the other language modalities. In T. V. Higgs (Ed.), *Teaching for proficiency, the organizing principle* (pp. 113-138). Lincolnwood, IL: National Textbook.

Lowe, P. Jr. (1985a). *ILR handbook on oral interview testing*. Washington, DC: DLI-ILR Joint Oral Interview Project.

Lowe, P. Jr. (1985b). The ILR Proficiency Scale as a synthesizing research principle: The view from the mountain. In C. J. James (Ed.), *Foreign language proficiency in the classroom and beyond* (pp. 9-53). Lincolnwood, IL: National Textbook.

Lowe, P. Jr. (1986). Proficiency: Panacea, framework, process. *Modern Language Journal, 70*, 391-397.

Savignon, S. V. (1985). Evaluation of communicative competence: The ACTFL Provisional Proficiency Guidelines. *Modern Language Journal, 69*, 129-134.

Valdman, A. (1987). The problem of the target model in proficiency-oriented language instruction. This volume.

DISCUSSION OF THE SOCIOLINGUISTIC ASPECTS OF COMMUNICATIVE ABILITY

John U. Wolff
Cornell University

I would like to address these questions from the point of view of the languages which I have been teaching: Indonesian, the national language of Indonesia, Javanese, a regional language of Indonesia which has a glorious past but a not-so-glorious present, Tagalog, which is a national language of the Philippines but has not quite made it to the top status-wise, and Cebuano, which is a regional language spoken by a large population but despised as the jargon of an uneducated peasantry. These languages are for the most part studied by people who have a career interest in some aspect of culture or development in Indonesia or the Philippines and who need these languages in order to carry out their career goals. They must have a first hand understanding of the ethnology of communication as a key to an understanding what makes the members of the speech community tick and they need to be able to interact with members of the community in a way that resembles the interaction of a full-fledged middle class member of that community. If we aim to teach our students how to speak like educated Indonesians or Filipinos that puts on us the burden of teaching them how to act and interact like Filipinos or Indonesians -- socially as well as grammatically. Stylistic or register variation is an important part of the paraphernalia which these speech communities have at their disposal to put their ethics of communication into action. It is this that I will address first.

From the learner's point of view, stylistic variation or variation based on social factors offers the most serious kinds of difficulty. Their usage is among the recalcitrant parts of a language. There are languages such as English, French, German, and so forth where the forms, the use of which entails a powerful speech act can be withheld from the student -- i.e., inappropriate, vulgar or profane forms are not going to be part of the curriculum and the student is unlikely to commit an absolutely unforgiveable breach of social rules unawares. The only minor problem for German is the use of du and sie, which have to be introduced to the student in the earliest stages, but even there the rules for when to use which are not complicated; I mean they do not really involve the assimilation of a whole new ethic or concept of how the world goes round, and an error or two is not going to become a source of strong tensions. There are other languages in which almost every time one opens one's mouth, a strong statement is made about human relations, status, or what kind of person everyone involved in the conversation is. These are languages whose cultures revolve around human relations and where each breach of linguistics etiquette has the potential of creating serious tensions. Japanese is probably a language of this sort, and of the languages which I have taught, Javanese from the island of Java belongs to this group. We will come back to this shortly.

Let me say a couple of words about the Philippine situation. There is a third group of languages, which are used by the educated native speaker only under very limited circumstances. They are used only in the family or perhaps to people very much in the inner circle. Outside of these small circles, they are confined to noneducated people in the community. This was formerly true of Tagalog and to some extent probably still is. It is certainly true of Cebuano. As soon as a foreigner uses one of these languages, he makes a statement that his interlocutor is an intimate, or if that is patently not the case, it is a statement that the interlocutor is uneducated and cannot be talked to in the appropriate language. In these cultures, we as foreigners have difficulty using the language. By virtue of being foreigners we cannot present ourselves as uneducated people who are forced to use the language in question, and if we persist with this language, the only possible conclusion is that we view the addressee as an uneducated native. This is a problem for which there is no satisfactory solution. The students have to be made aware of it at a fairly early stage and must be taught skills in creating fictions whereby they can define themselves as intimates.

Let us now go back to the Japanese-Javanese problem. The Japanese-Javanese problem is amenable to solution, but is immensely complex for the student to deal with. In Javanese, from the beginning, the student has to learn two and possibly three and four variants for a good portion of the basic vocabulary. Even if the forms used in conversation with intimates are withheld for the first year of instruction, the student immediately must learn to distinguish the varying forms which refer to personal activities and possessions and the relative status which each of these ascribe. To understand the *weltanschauung* which underlies the rules which tell how these various forms are chosen is no small potatoes. As if these problems were not enough, the social rules demand a certain amount of indirection, circumlocution and reference by inference; and failure to adhere to them makes a strong statement about social relations. Fortunately, one can get far in most of the Javanese areas without knowing Javanese: not only is Indonesian widely spoken in educated circles, but a repetition of a few memorizable formulas is sufficient to set the scene, give the impression that one is plugged into the Javanese *weltanschauung*, and give everybody concerned the proper status. On the other hand and unfortunately, Indonesian has taken over some of the Javanese complexity. Indonesian is dominated by the Javanese, and the Javanese ethnology of communication or method of interaction has become the national norm, such that some of the sociolinguistic complexity has spilled over into Indonesian. Thank heavens, not much of the rich treasury of alternative forms has been taken over, and the vocabulary involved is fairly small and readily learnable. However, alas, the complexity of the rules has indeed been taken over. As I say, the Javanese predominate, and they carry with them to the wider national context their *weltanschauung* and their notions of how interaction should be carried out. Other non-Javanese Indonesians must learn the rules which arise from these notions if they are to survive on the national scene. There are two matters which the student must confront from day one: first is the system of address and, to some extent, also first and third person reference. Second, are the unspoken postulates which underlie all exchanges. First the pronouns: there are five or six forms of high frequency which refer to the second person "you". The word for "you" may also be dropped under appropriate grammatical circumstances. It makes a strong statement about the position of the speaker and the interlocutor vis-a-vis each other and in the totality of the social order if one chooses one alternative or another or if one chooses to construct the sentence so that it contains no form for "you" -- i.e., the choice of the avoidance pattern. The ethics of interaction demand that these statements be iterated and often redefined several times in the course of the conversation. A switch from one alternative to another is not unusual or indicative of a moment of social crisis as for example the switch from *ty* to *vy* or vice versa in Nineteenth Century Russia nor is it a once-in-a-lifetime thing like in German. Also the conversation must be more or less liberally sprinkled with titles as vocatives, to keep this all-important element of Javanese (and now national Indonesian) ethics at the forefront. This together with fairly complicated rules on how to ask questions, how to make requests, how and when to ask for pardon, express thanks, and engage in several other types of speech acts which in Javanese society are very powerful indicators of social structure, combine to make a formidable task out of the task of bringing foreigners to act Indonesian linguistically.

There are the other problems of stylistic variation which have their counterparts in our Western languages. I can mention them only briefly. First, there are forms which indicate attitudes and presuppositions and clarify the context. Then there are the colloquial or dialectal forms which are part of everyday interaction and which one needs if one is to participate fully in communication -- if one is to behave normally. Certainly in Indonesian as much as in European languages one must observe propriety of speech by which I mean the register choice of colloquial or noncolloquial variants, which has an important role in defining the self and is also a statement of relations. A too formal form in a context in which a certain degree of intimacy or friendliness is required creates barriers and can thwart the goals which the foreigner is trying to achieve. A too colloquial form may make a statement of social relations which causes raw feelings and likewise thwarts goals. Colloquial forms cannot be withheld from the student. The student must know them to interact properly. Further, the student must have a thorough-going knowledge of colloquial forms in order to understand the contemporary written language, which makes heavy use of colloquialisms and dialectal forms for communicating between the lines. On the other hand, introduction of these forms into the instructional program leads the students to run the risk of using colloquial forms in inappropriate contexts. In the Philippines this problem does not confront the learner. These languages are in any case reserved for contexts in which the most informal variants are appropriate, and are not used in situations which require learned or formal styles of speech. For the foreigner English or English-Philippine mixture is the appropriate code.

Summary of the discussion session on SOCIOLINGUISTIC ASPECTS OF COMMUNICATIVE ABILITY: THE ISSUE OF ACCURACY VS. FLUENCY

Lisa Harshbarger
Indiana University

Albert Valdman opened the discussion by responding to a question raised by a respondent to his paper: Given that a classroom environment does not readily promote students' simultaneous achievement of all aspects of the pedagogical norm, which aspects should be made teaching priorities? Valdman selected actual native speaker behavior and acquisitional constraints as teaching priorities, given the little that is known about all of the criteria. He also indicated that the language classroom should present authentic texts across a wide variety of social matrices, in order to give the learner pedagogical criteria that reflect actual language use and encourage appropriate learner response, in terms of structure, sociolinguistic and discourse criteria, and effectiveness. There was some discussion of the role of learner error in this scenario, and the group concluded that neither grammatical nor sociolinguistic error was to be encouraged, but treated as a topic of classroom discussion.

A participant then noted that the representation of learner proficiency in the guidelines outlined by Pardee Lowe indicated that the development of language proficiency occurs in a hierarchical scalar fashion. It was argued that this formulation of the process of second language acquisition includes three problematic assumptions. The first assumption is that second language acquisition is linear and can be placed in a learning hierarchy, when research tends to show that it is not. The second assumption is that the behavioral descriptors of a particular level of proficiency correspond to some sort of set of linguistic universals, when researchers in the field of second language acquisition have yet to define such universals. The final problem with this model of proficiency is that it does not represent learner linguistic development as occurring within wide parameters of variability. Second language acquisition was defined by this speaker as a progression of individual function variables that grow exponentially, and he was concerned that representing such a process as if it were linear would encourage teachers to design their curriculum in that way, rather than incorporating more appropriate teaching techniques that focus upon learner variability.

Pardee Lowe then responded to these observations by noting that the proficiency test used by the Defense Language Institute was developed out of a needs analysis of what language skills were essential for communication in certain positions of responsibility. Out of this needs analysis, a hierarchy of skills did evolve, based on the difficulty of the linguistic tasks performed sucessfully by the test subject. This was challenged by some participants, who claimed that the test in fact only indicated that a person who ranked low on the proficiency scale was placed there because he or she exhibited low-level behaviors as defined by the scale itself, and that individual teachers would have widely varying definitions of what constituted an easy or difficult linguistic task. In response, some discussants commented that the hierarchy of behaviors established by the test was implicational, in that one set of behaviors tended to preclude others. Other discussants reminded the group that the use of this test in an academic context came about because of the need for teacher accountability, and that the test may not be an indicator of what a learner actually can do. A concluding comment on this discussion was that it seemed to revolve around how to label the hierarchy that is present, rather than whether or not such a hierarchy exists.

There was a brief discussion of whether William Labov's description of the variations of native speaker speech actually stigmatized non-educated native speech or whether this negative view stems from the applications of his description by other researchers to their studies. This led to a broader discussion of the types of native speakers American students of foreign languages are likely to encounter and the linguistic requirements of such encounters. It was noted that items from standardized listening comprehension tests such as the TOEFL often are spoken in "TOEFL English" (highly standardized speech free of error or even normal hesitancy) because both the learners being tested and professional test developers expect this. This led the group to ask what precisely foreign language teachers want such a test to measure, and if the Oral Proficiency Interview was being used to evaluate too many different kinds of populations.

The issue of test validity now became a focus of the group's discussion. It was asked if "experiential validity" was the sole criteria by which claims of the OPI's validity were made. The response by testers from the Defense Language Institute to that question was that some psychometric validity tests were done in 1959, but never published, and that it was hoped that academics from this Symposium and other groups would develop an interest in conducting such studies. It was also pointed out that triangulation of testers through analysis of their protocols and personal interviews had been conducted at the Institute, and that all testers tended to use four references when conducting their interviews: the level definitions provided by the test guidelines, language behavior of educated native speakers, language behavior of non-educated native speakers, and language behavior of non-level 5 non-native speakers. In the final analysis, these respondents concluded, user satisfaction is the ultimate test of validity. If those individuals certified by the Oral Proficiency Interview as proficient in that language are capable of doing certain tasks in that language to the satisfaction of those who hire them to do those tasks, then the test is valid. Some discussants noted, in conclusion, that psychometricians could legitimately question whether or not the test was being used appropriately in all contexts, and that documentation of the process by which final protocols and interpretation of a test are established is considered essential in any description of that test.

The final minutes of the discussion focused upon whether or not the use of the educated native speaker as the criteria for oral proficiency in the ACTFL guidelines was an oversimplification of the nature of actual proficiency in a particular language. While one participant agreed that this notion of proficiency probably excluded important linguistic behavior that was significant, he also suggested that two important aspects of the guidelines needed to be considered. First of all, as one moves up the proficiency scale, the amount of linguistic material that test subjects are expected to control grows geometrically. Secondly, careful speech is a significant aspect of a high level of proficiency. Rather than considering the continuum of speech behavior as ranging between articulate and less articulate forms, it was pointed out that one could consider the continuum to range between planned and unplanned speech, with the possibility of both extremes of performance indicating a high degree of articulateness.

Overall, the discussion focused upon the relationship of any standardized proficiency measure to 1) authentic linguistic input in the target language to which a learner may be asked to respond, 2) current understandings of how the process of second language acquisition is behaviorally indicated (and the assumption that such indicators can be placed in a hierarchy), and 3) professionally developed definitions of test validity. Those advocating a broader use of the ACTFL guidelines for proficiency testing emphasized practical considerations of test development (such as user validity) as opposed to theoretical issues. Those questioning the appropriateness of the guidelines argued that the current proficiency measures do not correspond well to current understandings in the field of second language acquisition of the nature of the three areas listed above.

FEATURES OF PRAGMATIC AND SOCIOLINGUISTIC COMPETENCE IN THE ORAL PROFICIENCY INTERVIEW

Heidi Byrnes
Georgetown University

Summary: The paper examines the rating scale and the testing procedure of the oral interview from a discourse and sociolinguistic perspective. While the rating scale has generally been characterized in terms of function, content/context and accuracy, its underlying motivation derives from the sociolinguistic features of appropriateness and acceptability. Given the inability of any testing of oral communicative ability to capture the multifaceted pragmatics of language use, the paper suggests that a careful examination of the processing capabilities which underlie speakers' different ways of handling various types of meaning chunks or propositions might serve as a valid indicator of their pragmatic and sociolinguistic competence.

1. Introduction

The evaluation of speaking proficiency according to the ACTFL rating scale, its testing protocol, the socalled oral proficiency interview (OPI), and its procedure for assigning ratings to samples of speech have sparked a significant amount of controversy. One issue raised is that pragmatic or sociolinguistic parameters are inadequately represented while discrete-point factors receive a disproportionate weighting (Savignon, 1985).

The belief that ACTFL proficiency assessment places too much emphasis on the notion of accuracy, particularly grammatical accuracy, is understandable from two perspectives. It reflects a concern that the prevailing, largely prescriptive, static, and structuralist assumptions about L2 oral language use are likely to dominate in the OPI as well. At the same time it asserts the challenge inherent in any assessment of oral language, particularly through a single testing procedure.

As for assumptions about L2 spoken language, the pedagogical literature is far from providing extensive coverage for the dynamic, interactive, and negotiated quality of oral language. In linguistic literature, the investigation of oral language use pertains primarily to L1 language and clarifies particularly its relationship to the received norm of literate language (i.e. Ochs, 1979; Chafe, 1982; Tannen, 1982). In those studies where L2 spoken language is the focus, investigations tend to concentrate on specific speech acts (Cohen & Olshtain, 1981) and rarely look at such global performance features as hesitations, false starts, repetitions, and corrections. Consequently, their possible meaning for L2 learners' linguistic, pragmatic, and sociolinguistic competence, as the L2 is acquired over an extended period of time and used at different performance levels, is all but unexplored. In the same fashion descriptions of instructed L2 learners' sociointeractional ability, sociocultural ability and sociocognitive aptitude, critical components of sociolinguistic competence (Kramsch, 1986), are nearly totally absent.

These circumstances on the side of research and scholarship make plausible the second concern about the OPI, namely that raised from the testing perspective: given our limited knowledge of the mainfestations of sociolinguistic competence, except for an awareness of its dependency on a multitude of characteristics inherent in infinitely configured communicative situations, it seems totally inappropriate to assume that a single testing procedure and a single rating hierarchy can make valid, reliable, and generalizable statements about L2 speakers' communicative ability.

Thus it is not surprising that proficiency assessment has been compared with norm-referenced measurement of performance features in a behavioristic vein (Kramsch, 1986), as it claims a hierarchy of competence and a level of predictability that is at best unfounded.

Cognizant of these concerns I nevertheless suggest that we not dismiss spoken language assessment through the OPI prematurely but, as a first step, analyze the rating scale and the testing procedure from a discourse and sociolinguistic perspective. A simple defense of the status quo is neither intended nor likely to be desirable. But we may well find that the bases for the rating scale are not unsound and that their articulation with a sense of awareness of language use, L1 and L2, beyond the sentence level and its functioning in conversational exchanges (Brynes, 1987) is therefore a valid approach toward better understanding of instructed L2 learner language.

In other words, I find it necessary, first to be mindful of the tension between the theoretically ideal and the practically feasible, second to compare that against what the current interviewing procedure manages to accomplish (Canale & Swain, 1980), and, finally, I propose that the distinct gap in our awareness of L2 learner communicative ability can begin to be filled through the careful observation of communicative performance as it is available through the criterion-referenced (*not* norm-referenced) assessment of the OPI from a sociolinguistic perspective, a psycholinguistic perspective, and a structural perspective.

I will begin by addressing the extent and manner in which the current rating scale can and does incorporate pragmatic and sociolinguistic features. The result of this analysis will then be applied to the current testing procedure. I will close with an overall assessment of that procedure and some suggestions for the direction future discussion might profitably take.

2. Of hierarchies and norms

The evaluation of oral proficiency under the ACTFL system proceeds from the stance that there exist two complexly intertwined hierarchies, one with respect to types of speech events, and another with respect to different realizations of one and the same type of speech event. Positing such a hierarchy is not to be construed as an assumption of a simplistic ordering among speech events: in foreign language pedagogy the functional-notional perspective has amply demonstrated the impossibility of applying simple hierarchies to the social event of language use. Even so, any assessment of L2 language use must first refer to a formal linguistic repertoire which a user is capable of handling, although its real goal is to determine the functional extra-linguistic repertoire, understood in the ACTFL rating scale as task universals, which the user is likely to be able to perform with that degree of linguistic competence. There are noteworthy and well-known limitations to such a procedure, but, in the absence of other practicable and generalizable approaches, one should also acknowledge its significant potential for coming to grips, in a systematic way, with the social dimension of language use.

The two hierarchies mentioned previously can be thought of as vertical and horizontal axes respectively (ACTFL Proficiency Guidelines, 1986). The vertical hierarchy reflects what are taken as native speaker judgments about the increasingly more elaborate language ability required to perform the tasks that have been identified. In a short-hand fashion their progression can be characterized as extending from naming and listing, to stating, narrating, describing, supporting an opinion, and negotiating or playing out a complex abstract issue or stance.

The horizontal hierarchy which concerns differences in the realization of one and the same task does not primarily pertain to absence or presence of errors in the usual sense, as has been suggested. Instead, much more elusively and at the same time more pervasively, it takes recourse to the phenomenon of sociolinguistic appropriateness, to the perceived fit between what are taken to be the speaker's intentions, given the task that was performed through language, and the anticipated message form.

Thus, while both of these hierarchies have generally been characterized in terms of function, content/context and accuracy, their underlying motivation derives from the key sociolinguistic features of the appropriateness and acceptability of a given speech sample vis-a-`vis the norms of the speech community. It is imperative that norms be understood here as ranges of acceptable behaviors whose limits shift and adjust to numerous factors constitutive for a speech event, among them setting, topic, intended message, and, of particular interest in our case, interlocutors.

On the issue of norms, much confusion has arisen about the two endpoints of the ACTFL scale, the non-functional non-native speaker and the non-native speaker who is indistinguishable from an educated native speaker. The intent of the scale has never been to judge learners' complexly varied, restricted interlanguage against what educated native speakers would have done in the same situation, educated native speakers who themselves come with all kinds of language performance profiles. Instead, it is an indication of the sociolinguistic sensitiveness of the scale to allow for non-native speaker performance to be acceptable in many ways that fall short of native norms. In other words, there is such a thing as norms of language use which native speakers concede to and apply toward non-native speakers.

But that is only part of the matter; more interesting is the consideration that these norms, too, shift with the perceived level of the interlanguage observed, thus leading to an ordered set of different expectations and different norms (Faerch & Kasper, 1983). There have been first attempts to capture the special characteristics of those successive norm shifts (Higgs & Clifford, 1982). Unfortunately the interpretation of those attempts has seized upon the increased importance of grammatical accuracy in native-speaker judgments as being characteristic of the progression, and, by extension, of the entire ACTFL scale, totally disregarding that it is not grammatical accuracy *per se* that becomes paramount, but its constitutive role for meaning and the powerful sociolinguistic metamessage associated with accuracy. Indeed, what I take to be the intent of that term might be better served by the term "appropriateness". The metamessage carried by violating its tenets can be as straightforward as "the speaker does not have the linguistic resources to express what she or he wants to say and therefore troublesome misunderstandings arise or communicative ends are not achieved". Or it might be as judgmental and at the same time irresistible as "the speaker is an insensitive, inconsiderate outsider who does not respect the norms of behavior of the linguistic and cultural group with which I, as a native speaker of this language, identify myself and which I value very highly".

While there seem to be different tolerance levels of native speakers toward different features of non-native speech (Chaudron, 1983; d'Anglejean, 1983), one can say that this latter kind of judgment is much more likely to occur with non-native speakers who have already acquired a respectable command of the second language than with those who are obviously struggling with the form inventory. In other words, sociolinguistic features of acceptability become much more prominent at the higher end of the scale.

The charge by critics of the OPI of an overemphasis on grammar is perhaps in no small part a reflection of the fact that proficiency advocates themselves have not elaborated on the increased weight the scale accords the sociolinguistic component and have scarcely spelled out which qualities of language use are indicative of higher sociolinguistic competence. Even so, the descriptors used in the rating scale, among them the concepts of tailoring of language to the audience, of acceptability, and the awareness of interactive norms and discourse strategies, indicate a clear commitment to sociolinguistic norms. Since they are not crucial components in the special norms accorded beginning speakers they do not become prominent until the Advanced and Superior levels. Granted, references to sociolinguistic competence are less than prominent and presuppose an intuitive awareness of the features in question rather than providing elaborate research backing, a presupposition that is likely to prove false. But that should not be taken to mean that sociolinguistic issues are not part and parcel of the entire hierarchy and that they cannot be articulated within the current framework.

3. Overt and covert pragmatic features

I have already given a brief characterization of the task hierarchy. Let me expand on this aspect from the viewpoint of pragmatics or, if you will, discourse competence.

There is a way in which the ACTFL task hierarchy can be summarized by saying that the appropriate performance of the tasks is relatable to linguistic surface realizations at the word, sentence, paragraph, and discourse level, with the latter two, in particular, depending on various overt and convert devices of coherence and cohesion. While I would not dispute the appropriateness of that characterization as far as it goes, I would like to shift the perspective from what I think is essentially an emphasis on finished product to an emphasis on underlying dynamics. These dynamics have at least two major components, the psycholinguistic processing characteristics of speakers and the pragmatic stance speakers are able to take with respect to their message and their interlocutors. Both can be observed only indirectly but powerfully nonetheless, under the rubric of what kinds of meaning chunks speakers can express in what kinds of language chunks. I take this relationship to be an important component of discourse competence and of pragmatic competence, indirectly reflected in the rating procedure and the hierarchy. Its particular value for articulating the hierarchy into a viable frame of reference for speaking derives from its combining an internal developmental psycholinguistic progression with an external norm-oriented pragmatic dimension.

As already emphasized, when we say that there exist different levels of speaking ability we are really saying that there exists a certain fit, based on a set of expectations, between communicative tasks and their linguistic realization. However, instead of characterizing these relizations in terms of surface syntax it might be more insightful, for an awareness of the pragmatic component of the rating scale, to think of them in terms of ways speakers have of expressing propositions. The reason for this is that how propositions are handled reflects on the one hand psycholinguistic processing constraints under which second language speakers operate particularly strongly, on the other it captures the special qualities inherent in any speech production along with the undeniable variation among speakers in their ways of manipulating language. A scale that aims to be pragmatically sensitive must address both concerns.

In examining how propositions are handled we would find answers to questions such as these: Are propositions expressed primarily at the naming stage which is closely tied to the lexical level? Can they be processed in simple relational terms, the essence of basic syntax? Is there evidence for the ability to hold a number of propositions simultaneously and to arrange these in a variety of increasingly complex ways with respect to each other, among them foregrounding, backgrounding, coordinating, subordinating, juxtaposing, attributing, assigning logical or chronological relationships, anticipating, referring back, etc? Does the speaker exhibit the capability of selecting from among the major propositional relationships those that serve best her or his communicative intentions or the demands of the communicative task at hand?

At first sight one is tempted to reply that answering these questions shows little relationship to pragmatics as the term is commonly understood, that it would primarily address linguistic phenomena, or that its real value would be for pointing out retrieval difficulties of the learner. Or one might raise the concern that a distinction between speakers on the basis of how they realize propositions has more to do with their intellectual maturity than it does with their language competence. However, I would suggest that, while the interpretations mentioned are indeed appropriate, the progression implied by the previouis questions is also a pragmatic one, in so far as the speaker is gradually enlarging her or his linguistic repertoire along a number of pragmatic axes. Among these are axes that characterize language use as extending from situation-bound to open-ended, from reactive to initiating, from overtly interactive to covertly interactive, from involved to detached, from speaker-bound to interlocutor-attentive, from uninvested to invested, from highly predictable task-specificness norms to individualized realizations of language-specific norms, among them norms of genre, style, register, or aesthetic valuation (see also Galloway, 1986).

In other words, I would see a relationship between the kinds of semantic processing speakers reveal through their language use and their pragmatic competence in that language (Garrett, 1986). When we say that somebody is a competent speaker of the language we are, strictly speaking, not so much referring to the correctness and to a surface elaborateness she or he is able to mobilize for message transmission. We are really referring to a host of pragmatic features. Among them are:

-this speaker, when called for, is able to show, through language, an independence of the immediacy of the situation;

-she is able to take note of the wide variety of features inherent in a setting and seeks to make them responsive to her needs through language;

-he has a level of awareness of the expectations of the interlocutors that allows him to frame his message so as to be recreatable and therefore meaningful for the hearer;

-her message is formulated in such a fashion that various markers (hesitations, pauses, corrections, augmentations) and appropriate redundancies help the listener follow it easily;

-he continuously monitors listener comprehension signals, actively seeking them should they be insufficient, and adjusting his message accordingly;

-her use of discourse space is appropriate to the communicative task. Along with the ability to respond to the immediacy of communication, which usually results in lengthier and redundant speech, the speaker can also plan discourse, a fact that reveals itself in reduced redundancies;

-his message becomes compelling and dynamic, not merely because of overt give-and-take but because of covertly interactive quality that arises out of anticipating the hearer's communicative expectations;

-she is willing to take account of her audience, by delicately balancing positive and negative face (Goffman, 1967);

-he is prepared to invest his ego into a message, rather than distancing himself from it (Galloway, 1986).

In themselves none of the above-mentioned ways of expressing propositions carries a judgment about speaking proficiency; only in relation to global expectations for a wide range of communicative events, and with respect to refined expectations vis-`a-vis a specific communicative event can such judgments be made.

To give an example, it is perfectly appropriate to respond to a naming task by listing individual words. However, consider the following not at all infrequent occurrence with second language speakers. A narrative task is handled in terms of one or two initial minimal sentences. Interlocutor questions are required to elicit further minimal reactions, where each one arises from a highly targeted request by the listener that the narrator address a certain proposition which must be assumed to have some import for the story. Depending on the patience, kindness, and level of interest of the listener/questioner the story actually might end up being recreated by the listener. But we would here have an instance of the non-native speaker not following the accepted sociolinguistic rules of how a story is to be told. This could be due to any number of extra-linguistic factors, in themselves potentially weightier than the observed linguistic shortcomings. But in the absence of immediate analysis by the participants themselves, a procedure that in no way guarantees the correct answer, we must make do with performance data. If such performance data analysis is backed by extensive correlational evidence then assigning a lower rating to such a performance is really an indication of the importance of pragmatic and sociolinguistic components. To rephrase, while the rating scale may, to a large extent, be expressed in terms of surface linguistic features I believe its real intent is to tap pragmatic and sociolinguistic competence through those surface language features.

We come here to two crucial issues in the assessment of speaking ability. First, in professing its basic feasibility we must face up to the fact that while, for native speakers, linguistic options are the result of semantic options, in the performance of language learners, the reverse may be the case. Linguistic, psycholinguistic and pragmatic options may ultimately determine what can be meant and the way in which it can be meant (Canale & Swain, 1980; Bialystok, 1983).

Second, a rating scale will never be able to capture the reality of speech in all its multifacetedness (Bachman & Savignon, 1985). It will always, even if only tacitly, assume that evidence for performance feature x is grounds for extrapolating from it the speaker's ability to handle performance feature y. The key question is, what are the features that are considered baseline, and which features can, with some justification, be extrapolated? As it stands, the ACTFL rating scale represents the assumption that evidence for elaborated processing capabilities of the variety just described justifies the extrapolation to pragmatic and sociolinguistic performance features which neither a rating scale nor a testing procedure can ever hope to capture but which, in the end, are crucial for the appropriate functioning of the non-native speaker. This assumption, though certainly in dire need of further explication, nevertheless is one that finds support in interlanguage research (Bialystok, 1983).

4. Pragmatic and sociolinguistic competence in the oral interview

When we speak of tapping pragmatic and sociolinguistic competence we are immediately struck by the limitations of any testing procedure, even one that, like the oral interview, is relatively authentic in that it focuses directly on total language behavior rather than on its component parts as the majority of tests do. Nevertheless, it represents no more than a minute sample of the virtually unlimited possibilities for speech events.

There can be no doubt that the interview gives primary evidence for only one kind of oral language use, namely the response a second-language speaker is able to provide to a single individual, the tester, in a setting that is marked by a power-relationship, whose intent is precisely that of testing oral language use. Little seems conducive in this setting for a surfacing of the crucial interactional, negotiative quality of speech, management features can only be tapped in a rudimentary fashion, and the function of language as a social bonder or as the primary signaling system for the myriad of roles we play in our daily lives, to mention just a few aspects, simply cannot appreciably come into play.

However, even this restricted primary evidence can yield a host of information if interpreted under another subfeature of the ACTFL rating scale's component accuracy, namely fluency. Let us examine some pragmatic and socio-linguistic properties as they are generally summed up by the term fluency.

The first set broadly refers to characteristics of spoken discourse as contrasted with written discourse. Most important among these is the feature of spoken language being "unplanned" (Ochs, 1979) or, as Enkvist puts it in view of the fact that all language involves planning, though at different levels, "impromptu" (Enkvist, 1982.). This means that, under the time pressures inherent in speech production, advanced editing will be minimal, resulting in seemingly fragmented speech. The planning span will be such that idea units will be short and generally expressed in syntactically simple and lexically restricted fashion (Chafe, 1980). In addition, these idea units which are realized as intonation units will involve numerous hesitations, false starts, repetitions, etc. and will themselves be conjoined by simple chaining or coordinating. This means that, propositional density which requires planning is rather low.

In other words, if spoken language as a whole is compared with the literate norm it is characterized more by disfluencies rather than by fluency, although we use that attribution in describing some speakers while withholding it from others. It seems, then, that fluency is first of all a relative concept and secondly an interactive concept that depends just as much on the listeners' judgments as it does on the speaker's characteristics. Put another way, we can say that "fluent" speech is not so much flawless as it is disfluent in ways that are acceptable to the listener (Deese, 1980).

Given the crucial role of listener judgments in language assessment, information about such acceptable ways of being disfluent takes on a special significance. First studies exist to show how perceived fluency correlates with perceived higher levels of proficiency and what language features are considered crucial for both judgments (Schwartz, 1980; Gaskill, 1980; Stewner-Manzanares, 1983; Kasper, 1985; Salo-Lee, 1987). It turns out that the ability to use various repair strategies, among them hesitations, restarts, clarifications, lexical searches, and correction invitations, increases with the level of proficiency, and is directly related to an assessment of greater fluency. In other words, these background repair strategies, as distinct from simple linguistic repairs, which we intuitively associate with lower levels of competency, are actually indicators of a higher proficiency. They serve both continuity, including the ability to fill the available discourse space, as well as discourse organizational purposes.

While the summary descriptions of the major levels which make up the rating scale do not refer to such manifestations of fluency the progression underlying the scale as well as actual rating practice clearly take them into account. For example, the ability to produce paragraph-length speech, criterial for the Advanced level, very often is manifested not in propositionally dense, lexically precise and organizationally tight polished paragraphs as we associate them with literate language but rather with the extensive use of the kinds of background repair strategies referred to above. Likewise, when the Intermediate level speaker is characterized as essentially reactive it is in no small part because of a tendency to look for other-initiated and other-completed repairs, or other-initiated self-completed repairs, even though current interviewing protocol disallows such behavior on the part of the tester.

Thus, while fluency has yet to be spelled out in terms of sociolinguistic features of appropriateness the current rating scale would readily seem to allow such an emendation.

5. The pragmatics of testing

We have repeatedly referred to the interactive quality not only of the communicative event itself but also of its assessment. Of necessity that realization puts a significant burden on the tester-rater, a burden that is made even more critical due to the highly restricted primary evidence that is obtained in the OPI. Nevertheless, I believe that, in the hands of a skillfull tester, secondary evidence can become so powerful as to reduce the obvious pragmatic shortcomings of the setting dramatically, and, in fact, permit some valid inferences.

To me, among many components that can be isolated the key factor in such skillful testing is the centrality of message transmission, of language arising out of a desire to communicate real meanings, rather than contriving occasions that would allow for mentally checking off whether the candidate can handle structure x or has available lexical item y. Strictly speaking neither one of these can be a basis for assigning a rating. They are only a small part of the surface evidence we need for the underlying processing capabilities which, in turn, reveal pragmatic and sociolinguistic capabilities.

With a keen awareness on the part of the tester of what constitutes natural conversational behavior, including natural listening behavior, curbing it only in so far as it would preclude an optimal and efficient sampling of speech on the part of the candidate, I have observed over and over again how the dynamics of message transmission become so powerful in themselves that the candidate can present an amazingly multi-faceted performance, even at the lower level of language use (Byrnes, 1986). And once beyond the initial hurdles, the imperatives of message transmission, coupled with the speaker's greater sociolinguistic competence, allow the restrictiveness of the setting to move further and further into the background and a true conversational exchange to take place. Indeed, as mentioned before, it is the hallmark of such speakers to be able to leave the constraints of the here and now and with their language create and define their own conversational parameters within a given setting.

Conversely, when the tester does not bring to the task that level of interpersonal involvement and willingness to become a true conversational partner within the constraints of the event then little of the kinds of evidence for pragmatic competence I have mentioned can surface. On a philosophical level, this is really what an unratable sample is all about, namely the tester's inability to draw out of the candidate the best possible performance in a bias for the best, and inability that is directly tied to a disregard of the pragmatic dynamics of conversational exchanges and of oral language use in general. One might interject that, at the lower levels, candidates are unlikely to be able to muster much pragmatic finesse anyhow, and that at the higher levels they are quite capable of taking over. However, it is the in-between ground that is likely to suffer most under an inept tester, and, unfortunately, much of academic testing takes place here.

To summarize, in order to permit the clearest evidence for unperturbed language use a genuine conversational exchange, deemed the archetypical occurrence of spoken language, is crucial for offsetting the substantial constraints of the testing procedure.

6. Outlook

Where are we to go from here? As indicated from my characterization of the rating scale, I would consider these points to be in need of further consideration:

(1) The relationship between the primarily "linguistic" descriptors in the rating-scale and the pragmatic and sociolinguistic paramenters that underlie native-speaker judgments about non-native speakers, particularly at the more advanced performance levels (see also Albrechtsen et al., 1980).

(2) The degree of acceptance, or tolerance if you will, non-native speakers extend to non-native speech. In contrast with discrete-point studies that have been undertaken, global assessments seem most unstructive, particularly if they are substantiated by an inventory and, if possible, a ranking of the criteria employed. On that basis a greater awareness of the shifting norms for the appropriateness of non-native speech could be obtained.

(3) Across the spectrum of second language research the psycholinguistic and meaning-derived dynamics of language acquisition and language use must come even more into the foreground (for an example see (Kramsch, 1983; Edmondson, 1985). As teachers, it is our perception of the second language acquisition process which, for better or worse, is manifested in the language we present to learners and, consequently, the language learners learn.

While it may seem presumptuous to suggest that a controversial topic such as the proficiency rating scale and the evidence about speaking which its testing procedure, the oral interview, offer can have an important role in this set of inquiries I would, in the absence of other viable contenders, nevertheless strongly urge its consideration.

References

American Council on the Teaching of Foreign Languages. (1987). ACTFL proficiency guidelines 1986. In H. Byrnes & M. Canale (Eds.), *Defining and developing proficiency: Guidelines, implementations and concepts* (pp. 15-24). Lincolnwood, IL: National Textbook.

Albrechtsen, D., Henriksen, B., & Faerch, C. (1980). Native speaker reactions to learners' spoken interlanguage. *Language Learning, 30*, 365-396.

Bachman, L., & Savignon, S. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. *Modern Language Journal, 70*, 380-390.

Bialystok, E. (1983). Some factors in the selection and implementation of communication strategies. In C. Farerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 100-118). New York: Longman.

Byrnes, H. (1987 in press). Speech as process. *Foreign Language Annals.*

Byrnes, H. (1986). Second language acquisition: Insights from a proficiency orientation. In H. Byrnes & M. Canal (Eds.), *Defining and developing proficiency: Guidelines, implementations and concepts* (pp. 107-131). Lincolnwood, IL: National Textbook.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*, 1-47.

Chafe, W. (1980). Some reasons for hesitating. In H.W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (pp. 169-180). The Hague: Mouton.

Chafe, W. (1982). Integration and involvement in speaking, writing, and oral literature. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy* (pp. 35-53). Norwood, NJ: Ablex.

Chaudron, C. (1983). Research on metalinguistic judgments: A review of theory, methods, and results. *Language Learning, 33*, 343-377.

Clark, J.L.D. (Ed.). (1978). *Direct testing of speaking proficiency: Theory and application.* Princeton, NJ: Educational Testing Service.

Cohen, A.D., & Olshtain, E. (1981). Developing a measure of sociocultural competence: The case of apology. *Language Learning, 31*, 113-134.

d'Anglejean, A. (1983). Introduction to special issue on native speaker reactions to learner speech. *Studies in Second Language Acquisition, 5(2)*, vii-ix.

Deese, J. (1980). Pauses, prosody, and the demands of production in language. In H.W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 69-84). The Hague: Mouton.

Edmondson, W.J. (1985). Discourse worlds in the classroom and foreign language learning. *Studies in Second Language Acquisition, 7*, 159-168.

Enkvist, N.E. (Ed.). (1982). *Impromptu speech: A symposium.* Abo: The Research Institute of the Abo Akedemi Foundation.

Faerch, C., & Kasper, G. (Eds.). (1983). *Strategies in interlanguage communication.* New York: Longman.

Faerch, C., & Kasper, G. (1983). Plans and strategies in foreign language communication. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 20-60). New York: Longman.

Galloway, V. (1986). From defining to developing proficiency: A look at the decisions. In H. Byrnes & M. Canale (Eds.), *Defining and developing proficiency: Guidelines,* *implementations and concepts* (pp. 25-73). Lincolnwood, IL: National Textbook.

Garrett, N. (1986). The problem with grammar: What kind can the language learner use? *Modern Language Journal, 70*, 133-148.

Gaskill, W.H. (1980). Correction in native speaker-nonnative speaker conversation. In D. Larsen-Freeman (Ed.), *Discourse analysis in second language research* (pp. 125-137). Rowley, MA: Newbury House.

Givon, T. (1979). *On understanding grammar*. New York: Academic Press.

Goffman, E. (1967). *Interactional ritual*. Garden City, NY: Doubleday.

Higgs, T.V., & Clifford, R. (1982). The push toward communication. In T.V. Higgs (Ed.), *Curriculum, competence, and the foreign language teacher* (pp.57-79). Lincolnwood, IL: National Textbook.

Kasper, G. (1985). Repair in foreign language teaching. *Studies in Second Language Acquisition, 7*, 200-215.

Kramsch, C.J. (1983). Discourse function of grammar rules: Topic construction in German. *Modern Language Journal, 67*, 13-22.

Kramsch, C.J. (1986). Response to H. Byrnes' paper, presented at the ACTFL workshop Research on second-language learning and the ACTFL spoken proficiency test. Dallas, TX.

Ochs, E. (1979). Planned and unplanned discourse. In T. Givon (Ed.), *Discourse and syntax* (pp. 51-80). New York: Academic Press.

Salo-Lee, L. (in preparation). *Investigating the repair strategies of learners of German at different levels of oral proficiency*. Unpublished doctoral dissertation. Georgetown University, Washington, DC.

Savignon, S.J. (1985). Evaluation of communicative competence: The ACTFL Provisional Proficiency Guidelines. *Modern Language Journal, 69*, 129-134.

Schwartz, J. (1980). The negotiation of meaning: Repair in conversations between second language learners of English. In D. Larsen-Freeman (Ed.), *Discourse analysis in second language research* (pp. 138-153). Rowley, MA: Newbury House.

Stewner-Manzanares, G. (1983). *Turn-taking structure of second language learners of Spanish at different oral proficiency levels*. Unpublished doctoral dissertation. Georgetown University. Washington, D.C.

Tannen, D. (Ed.). (1982). *Spoken and written language: Exploring orality and literacy*. Norwood, NJ: Ablex.

COMMENTS

Dell Hymes
University of Pennsylvania

I am glad to have the opportunity to take part in this symposium, and want to thank Albert Valdman for making it possible. Bloomington is where I did my graduate work and met my wife, and where twenty years ago I returned for a conference at which Roman Jakobson spoke on 'Linguistics and Poetics', presenting a model of the speech event that was a starting point for my own efforts to develop the ethnography of speaking.

The symposium demonstrates beautifully the sometimes emergent character of rules of speaking. We have quickly established that the opening slot of a formal turn has two requirements: not 'Hello', or a term of address, but a joke, and the statement, 'I am not a language teacher'.

The task of saying something useful does indeed strain my strategic competence to the utmost. I know little about the state of research in oral proficiency. There are some five comments I should like to offer from the vantage point of a general concern with pragmatic and sociolinguistic competence.

First, Dr. Byrnes' concern for the training of apt testers, her awareness that the tester is part of the test, is very much in keeping with a growing concern for the interactive nature of performance and the production of verbal data.

Second, Dr. Byrnes' frequent pairing of the terms 'pragmatic and sociolinguistic' suggests that neither will do alone, although each is taken as wholly comprehensive by some users. Dr. Byrnes is sensitive, I think, to vagueness and overlapping in current attempts to plot the domain of communicative ability. That situation will not soon be resolved. We are in the midst of widespread interdisciplinary, or transdisciplinary, concern with language, a concern that finds expression in the widespread use of such terms as 'pragmatic', 'sociolinguistic', 'semiotic', 'discourse', 'communication', any one of which might serve alone. In linguistics there are attempts to define structural levels, mainly as between 'semantics' and whatever is beyond semantics. Often enough, however, these attempts leave unidentified something that is beyond whatever is beyond, falling short of the full scope of sociocultural organization of use of language. Attempts in terms of levels of linguistic structure must always fall short, perhaps, for two reasons. The notion of level is associated with the notion of total accountability of representation, the possibility of saying of all segments, phonologically, morphologically, syntactically, the unit to which each pertains. It does not appear that this can be done in terms of currently available analyses of speech acts and conversational principles. Less ethnocentric conceptions would help, but there remains the task of dealing with style. Communicative ability consists in an important part of control of a repertoire of styles, and styles may consist of the selecting and grouping together of features from several levels. Communicative ability is fundamentally a matter, not of any one level of language or sphere of language use, but of ways of speaking. In linguistics and the various frames of reference concerned with language there is still relatively little sustained effort to identify and systematically analyze the range of human ways of speaking as such.

To say this is to say that despite the availablity of comprehensive terms, such as 'pragmatics' and 'sociolinguistics', we still have a step or two to take, going beyond the features and relations we are able to analyze now, and seeing the further ways in which they cohere in use. In the slow growth of empirical adequacy in knowledge, through linguistics, of language as human ability, the initial step, from phonetics to phonology, and the major subsequent steps, to syntax, and to pragmatics, have required recognition of a level of functional relevance that brought to light features and relations not previously recognized. So it seems to me with our knowledge of ways of speaking. We seem to be largely prisoners of attempts to impose external classifications and conceptions of communicative activity, and to suffer from an absence of concern for building a sufficient empirical base. Phonologists and grammarians have a wealth of detailed data from around the world in terms of which to test and refine their models. Not so students of ways of speaking. We are hardly able to begin to speak systematically of what can count as 'formal' vs. 'informal', 'direct' vs. 'indirect', 'loquacious' vs. 'taciturn', 'relevant' vs. 'irrelevant', 'clear' vs. 'unclear', 'sincere' vs. 'insincere', or even to say what dimensions are required. Phonology and grammar transcended Europocentric frameworks and became adequately general through description of cases from throughout the world. Many theorists of speech acts, pragmatics, conversation and discourse seem content to stay at home.

In this regard--to make a third point--it is very good to hear of Modern Standard Arabic and Tok Pisin, and of the need to go beyond a naive conception of the 'native' speaker. I have been disturbed by the tendency in such research literature as I have seen, and the work of some graduate students, to rely on a simple dichotomy between 'native speakers' and 'non-native speakers'. A sociologist friend of mine once said that he always introduced class as a variable in any study, because, whatever else he hoped to find, differences according to class would always turn up. Differences according to native vs. non-native speaker perhaps will also always turn up, but whether or not they are what they appear is questionable. 'Non-native speaker', after all, is a wastebasket category. One can hardly assume that having English as a second (or later) language makes the whole world kin, alike in behavior and with non-native status as the explanation of the likeness. 'Native speaker' itself, indeed, is a small wastebasket, as if the fact that the language is the same makes the use the same. One could hardly be further from sociolinguistics and the principle that rules of grammar and rules of use are distinct, that communicative ability is socially constituted (Hymes, 1974).

The assumption that chronology and capacity are in a one-to-one relation was indeed a half-century ago an unrecognized limitation of the ideas of Sapir and Whorf as to the relation of language to culture and world view (cf. Hymes, 1966). They were able to write as if a first language was an only language (and therefore the instrument of every use of language). But of course, before one can determine the significance of a language for thought or behavior, one must determine its role within the life of a person and within a community verbal repertoire.

The terms 'native' and 'natural' are very popular and often used without thought for what may be implied as to a conception of 'nature'. From the evidence of Modern Standard Arabic and Tok Pisin, and on general theoretical grounds, it would seem desirable to drop the 'N' from designations such as 'NS' and 'NNS', and in cases such as these to speak simply of 'ES'--'Educated Speaker'.

The evident need not to impose inadequate preconceptions as to what is 'native' and 'natural' leads to a fourth point. Much of the discussion of foreign language proficiency testing understandably brings one or another theoretical perspective to bear on the practical work involved. The relation between theory and practice need not and should not be one-sided. There is an opportunity for practitioners to contribute to theory. We need to build theory on an adequate range of cases and practitioners can make a critical difference. Certainly they should read theory that exists with critical eye, never taking it for granted. This point seems worth stressing. We seem to be in a period in which proposals of a system of turn-taking, say, or of taking conversational implicature, and the like, can be read by students as if they were the distillation of years of systematic observation akin to Darwin's. Even in a course devoted to teaching that conceptions of language and its use are always socio-historically relative, affected by our own backgrounds, and that field work must seek to discover the features and relations that hold for the people in the particular situation, the first drafts of some field work papers will not start with the situation observed, but uncritically cite such papers as if normative for the English-speaking world and beyond. It is vital for practitioners to take into account the empirical base of any theoretical proposal--Los Angeles County, perhaps, friends and neighbors of an intellectual class, whatever--and not assume that everything is the same in Milwaukee, or among highway maintenance crews. The practitioner can both avoid ill-fitting theory and help change theory to fit.

As a fifth point, I would like to mention some work in discourse that has not yet become widely known. Much of communicative ability, especially at the upper levels of proficiency perhaps, involves the telling of stories. True ability consists not in simply saying how a story goes, but in performing it. In recent years it has become increasingly clear that oral performances of stories do not organize them in terms of sentences and paragraphs, but in terms of lines and groups of lines. A variety of features may enter into marking the lines and their relationships--intonation contours or tone groups, syntactic particles such as 'so', 'well', "then', etc., changes of tense and aspect, quoted speech, etc. The grouping of lines in relation to each other seems to be done in accordance with one or another of a very few alternative models. It begins to appear that linguistic theory itself must take into account as fundamental to innate competence, not only sentences, but lines and certain ways of grouping lines. Students of proficiency may gain increased precision in understanding the features and relationships that enter into narrative accomplishment in a culture, and the ways in which shared stories may be shaped to convey individual intent. (cf. Hymes, 1985, 1988).

Author Note

While on the topic of designation by initial, I could not help try to place myself in relation to those who study Languages for Special Purposes and other such. Someone like myself, who devotes himself to an obsolescent American Indian language, might be said to be concerned with a LULU (Languages Unnoticed, Little Understood).

References

Hymes, D. (1966). Two types of linguistic relativity: Some examples from American Indian ethnography. In W. Bright (Ed.), *Sociolinguistics* (pp. 114-158). The Hague: Mouton.

Hymes, D. (1974). *Foundations in sociolinguistics*. Philadelphia: University of Pennsylvania Press.

Hymes, D. (1985). Some subtleties of measured verse. In J. I. Hesch (Ed.), *Proceedings 1985* (pp. 13-57). Buffalo, NY: The Niagara Linguistics Society.

Hymes, D. (1988). Linguistic anthropology as ethnopoetics. To appear in a volume edited by S. W. Mintz.

RESPONSE TO HEIDI BYRNES

Claire Kramsch
Massachusetts Institute of Technology

Heidi Byrnes' paper is a timely reminder of what communicative proficiency is about, beyond grammatical accuracy and functional competence. She spells out correctly the overt and covert pragmatic features of speech that enable a NNS to behave appropriately in the foreign language, apply the appropriate discourse strategies and meet the paper interactional expectations. She stresses quite rightly that sociolinguistic competence is more than the mere use of an appropriate gambit: it means tailoring one's speech to one's audience, showing responsiveness to the needs and expectations of the hearer, investing one's attention and one's interest in the topic at hand "in a compelling, dynamic discourse." In short, Heidi Byrnes refers to what Canale and Swain (1980) and Sandra Savignon (1983) have defined as, on the one hand, sociolinguistic competence, or the ability to understand and produce language that is appropriate to the topic, purpose and social status of the participants, and, on the other hand, discourse competence, or the ability to combine and connect utterances in conversation to form a coherent and cohesive spoken text.

Byrnes then makes clear that it is the intention of the ACTFL scale to account for such sociolinguistic abilities. She further suggests that a good Oral Proficiency Interview can indeed assess sociolinguistic competence, if the scale is refined to accommodate the components of sociolinguistic competence, if attention is paid to the surface features of speech, indicative of communication strategies, and if the interviewer shows enough ego involvement to become a real conversation partner and not just a tester.

I see three problems with this view.

1. Sociolinguistic competence is an interactional competence.

It cannot be derived from the observation of isolated surface features of one individual's speech at the sentence level. It is a modular proficiency that is determined by the situation, the topic and the social status of both participants, as they attempt to understand each other's intentions and expectations. This is true of the testee and the tester. For both, native speaker and non-native speaker alike, the linguistic resources determine and are determined by the semantic options available and the interactional consensus that both strive to establish. The communication strategies they use to construct a common conversational reality between them is occasionally realized through linguistic structures; but it is one thing for psycholinguists to hypothesize from surface features of speech to psychological processes for research purposes, it is quite another matter to use these linguistic realizations or lack thereof as evidence of underlying communicative competence or incompetence for rating purposes. For example, let us suppose the tester wants to know whether the testee knows how to realize in French the function "responding to compliments" and (s)he says: "Oh, quel joli pullover!" If the testee responds nothing or "merci", it might be that: (1) the student has not recognized the intended speech act (in an exam situation such a compliment might be perceived as incongruous); (2) the student has perceived it to be another speech act, e.g., a personal gift, or the recognition of a deserved attribute, as is customary in American culture; (3)

the student doesn't know the culturally appropriate linguistic realization of the intended speech act in French, which should be something like "oui, vraiment? Il est déjà vieux!" The assessment of appropriacy will depend on the testee's perception of the tester's intention, which might itself be ambivalent in a testing context. This example shows how difficult it is to assess through an Oral Proficiency Interview whether the student actually knows or doesn't know how to respond to compliments in French.

The same caution has to be applied when taking certain communicative strategies as evidence of native-like fluency. Faerch and Kasper (1983) have made the distinction between reduction strategies (e.g. message abandonment, hesitations) and achievement strategies (e.g. restructuring, approximations and appeals for assistance). Without going as far as Omaggio, who warns against the use of any communicative strategy (1986, p. 290) we do have to differentiate between those strategies that avoid difficulties in communication and those that compensate for them. As Albrechtsen, Henriksen and Faerch admit, "we need to know more about the communicative potential of different types of communication strategies before we can advise learners on which strategies to adopt and which to avoid" (1980, p. 395).

2. Sociolinguistic competence is a sociocultural ability that cannot be measured on a norm-referenced scale.

Byrnes points rightly to the ambivalence of the word norm in this context. If we adopt a pedagogic variable norm, as she suggests, where are the cut-off points? If "merci" is rarely acceptable in response to a compliment in French, but we accept it at the lower ends of the scale, at which level will it cease to be pedagogically acceptable? Advanced learners have to deal with higher expectations, but then again they are more able and entitled to flout these expectations on purpose. Byrnes speaks of the necessity of obeying certain culturally determined, interactive norms. But these are shifting norms, that are class-, gender- and situation-specific; they are socially and politically determined, and have no hierarchy, no zero and no absolute points. Moreover, special norms accorded beginning speakers place the tester in the difficult position of having to lend the same benevolent ear to non-native speech as native speakers do, without exercising pedagogic indulgence.

Thus, the solution Byrnes proposes does not remove the dilemma. Trying to do as if there were no norms, by having the tester behave like a natural conversational partner makes the test more authentic, i.e., reliable, but by the same token much less generalizable, i.e., valid. Furthermore, the tester's ego involvement, advocated by Byrnes as a prerequisite for eliciting "interactional, dynamic and negotiated" oral language on the part of the learner, is not without problems. For one, "negotiation of meaning" is impossible to rate on the present ACTFL scale. Omaggio (1986, p. 5) even doubts that this aspect of communicative competence should be stressed at all: "Savignon's view of the notion of communicative competence emphasizes the negotiative nature of communication. This is certainly an important point, and one that should be kept in mind when assessing any given sample of speech, writing, or receptive skills performance. This view of communication competence does not lend itself easily, however, to determining in any precise way the degree to which an individual is communicatively competent." She concludes that, rather than teach communication strategies whose value is questionable, the teachers should concentrate on systematic error correction, especially at the suprasentential level of discourse. Secondly, the possibility of contesting the validity of the OPI on the grounds of insufficient "ego involvement" on the part of the tester opens the door to a litigiousness that is already all too prevalent in our society (Bachman, 1987).

3. Sociolinguistic competence includes general sociocognitive abilities.

The very skills mentioned by Heidi Byrnes: "independence of the immediacy of the situation", "awareness of the multiplicity of interactional needs and expectations in a communicative situation", "ability to tailor one's discourse to the expectations of one's interlocutor, independently of the overt give and take", all these are, in fact, decontextualization skills, that have in part to do with general metalinguistic ability to identify potential ambiguity and to conduct autonomous speech. Deborah Tannen distinguishes between oral and literate style in conversation, between contextualized and decontextualized features in oral discourse (Tannen, 1982). The first, or context-embedded skills, are grounded in the here-and-now of the communicative situation, they are dependent on collaboration in narration, and are related to what Cummins calls BICS, or Basic Interpersonal Communication Skills (Cummins, 1983). The second, or context-reduced skills, represent the ability to conduct an autonomous style of discourse, that preempts ambiguity, establishes coherence through rhetorical organization and shows evidence of abstract verbal reasoning. This second type of sociolinguistic competence has to do with Cummins' CALP, or Cognitive Academic Linguistic Potential.

Now, the more we refine the scale of sociolinguistic competence, the more we require evidence of CALP in conversational management. As we know, the display of one's CALP is determined by such factors as level of education (Bernstein, 1971; Scollon & Scollon, 1981), family background (Snow, 1987) and style of literacy (Brice Heath, 1986), ethnic conversational style (Tannen, 1982), personal choice (Thomas, 1983) and the interactional pressure caused, for example, by a testing situation. The dilemma is: the more we refine the scale, the more difficult it is to know what we are testing and the more discriminatory the OPI becomes against students from social and educational background with less decontextualized skills. If we are testing white mainstream literacy skills, then we should say so, if only for the sake of truth in advertising. But then, is the best way to test these skills a discrete point norm-referenced language test for specific tasks and functions?

In sum, sociolinguistic competence cannot be developed within a behavioristic frame work like that suggested by the ACTFL guidelines. Dell Hymes (1972), by criticizing Chomsky's notion of competence, did not intend to return to Skinner's notion of performance, but to expand the concept of language competence to include communicative competence. If sociolinguistic competence means resetting, in the light of the L2 experience, the values of social parameters of language use whose values had been set by the L1 experience (Chomsky, 1985), then we are dealing with more than a behavioristic training.

Sociolinguistic competence, as defined by Heidi Byrnes, is an extremely important goal in foreign language education, but it is still too ill-defined to be rated by the existing OPI. I have tried to show that it includes at least three abilities that are heavily dependent on extraneous factors: a sociointeractional ability, dependent on the joint construction of a "sphere of intersubjectivity" between two conversational partners (Wells et al., 1981, p. 47); a sociocultural ability, determined by shifting norms of social behavior; and educational background. As long as the criteria of sociolinguistic competence are not more clearly established, it can only be the object of formative testing, where "bias for best" is in order, not of evaluative procedures that claim horizontal predictability across contexts and a vertical hierarchy of norms.

Meanwhile, I would like to recommend pursuing the following areas of research:

- identify, as Heidi Byrnes suggests, the criteria of sociolinguistic competence on a higher level of abstraction, namely sociocultural parameters of discourse in a top-down approach, not discrete behavioristic objectives in the botton-up approach of traditional classroom and curricular practice (Lowe, 1986, p. 396).
- include these components in the rating scales without specifying any surface language features, situations or discrete functions.
- give separate ratings for sociolinguistic competence against the criteria chosen.
- accept and acknowledge the inevitable subjectivity of the ratings and find ways of counterbalancing it by either specifying the purposes of the test, or by offering opportunities for multiple interactional contexts with a variety of raters.

In conclusion: Byrnes recommends making the distinction between "the theoretically ideal and the practically feasible." Although she is right in reminding theoreticians about the realities of the classroom, such advice can be misunderstood. Focussing on the feasible has meant, for many advocates of the proficiency movement, applying learning styles and methodologies appropriate for the beginning levels to the intermediate and advanced stages. It has implied the belief that all learners at all levels use the same learning styles and strategies. This assumption is not only erroneous but precludes the educational development on which learning is predicated.

If we consider teaching, and the teaching of a foreign language and culture in particular, as a subversive art, then we need to keep alive the vision of the ideal cognitive, cultural and interactional growth necessary for the development of sociolinguistic competence. It is this vision, not present realities, that will enable us to elaborate the practically feasible in foreign language education.

References

Albrechtsen, D., Henriksen, B., & Faerch, C. (1980). Native speaker reactions to learners' spoken interlange. *Language Learning, 30*, 365-396.

Bachman, L. (1987). Problems in examining the validity of the ACTFL oral proficiency interview. This volume.

Bernstein, B. (1971). *Class, codes and control* (Vol. 1). London: Routledge and Kegan Paul.

Brice Heath, S. (1986). What no bedtime story means: Narrative skills at home and school. In B. B. Schieffelin & E. Ochs (Eds.), *Language socialization across cultures* (pp. 97-124). Cambridge: Cambridge University Press.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1-47.

Chomsky, N. (1985). *Knowledge of language: Its nature, origin and use*. New York: Praeger.

Cummins, J. (1983). Language proficiency and academic achievment. In J. W. Oller (Ed.), *Issues in language testing research* (pp. 108-129). Rowley, MA: Newbury House.

Faerch, C., & Kasper, G. (1983). Plans and strategies in foreign language communication. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 20-60). New York: Longman.

Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269-293). London.

Lowe, P. (1986). Proficiency: Panacea, framework, process? A reply to Kramsch, Schulz and, particularly, to Bachmann and Savignon. *Modern Language Journal, 70*, 391-397.

Omaggio, A. C. (1986). *Teaching language in context: Proficiency oriented instruction.* Boston: Heinle and Heinle.

Savignon, S. (1983). *Communicative competence: Theory and practice*. Reading, MA: Addison-Wesley.

Scollon, R., & Scollon, S. (1981). Narrative, literacy and face in interethnic communication. In R. O. Freedle (Ed.), *Advances in discourse processes* (Vol. 7). Norwood, NJ: Ablex.

Snow, C. (1987). Beyond conversation: Second learner's acquisition of description and explanation. In J. Lantolf & R. di Pietro (Eds.), *Second language acquisition in the classroom setting* (pp. 3-16). Norwood, NJ: Ablex.

Tannen, D. (1982). The oral-literate continuum in discourse. In R. O. Freedle (Ed.), *Spoken and written language* (pp. 1-16). Norwood, NJ: Ablex.

Tannen, D. (1982). Ethnic style in male-female conversation. In Gumperz, J. (Ed.), *Language and social identity*. Cambridge: Cambridge University Press.

Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistic, 4*, 91-112.

Wells, G. et al. (1981). *Learning through interaction: The study of language development.* Cambridge: Cambridge University Press.

Summary of the discussion session on DISCOURSE STRUCTURE AND COMMUNICATIVE STRATEGIES IN THE DEVELOPMENT OF COMMUNICATIVE ABILITY

Molly Wieland
Indiana University

The discussion began with a request for clarification in the use of the terms accuracy and appropriateness, a distinction that is central to the evaluation of communicative ability. One of the participants expressed concern that the term accuracy was being replaced by the term appropriateness in the Guidelines. It was felt that both accuracy and appropriateness have a place in language proficiency, and that calling accuracy appropriateness confuses the issue.

It was subsequently explained that the increased use of the term appropriateness was a reaction to a debate that had been going on over the past few years between proficiency people (or insiders) and outsiders (this distinction is discussed in detail later in this session). The outsiders believed that accuracy was being interpreted by the insiders as involving only grammatical accuracy, although there were some insiders who were using the term accuracy to refer not only to grammatical accuracy, but also to sociolinguistic accuracy. The adoption of the term appropriateness to refer to sociolinguistic accuracy was an attempt to distinguish the two types of accuracy.

However, another participant took a different point of view, saying that in the earlier versions of the Guidelines, the term accuracy was indeed used to refer specifically to language--specific syntactic and morphological features. The increased use of the term appropriateness in the 1986 Guidelines thus reflects not a clarification in the meaning of the term accuracy, but a broadening of the notion of proficiency to include sociolinguistic features.

It was agreed that the two terms are necessary for a complete description of language proficiency. What remained was the task of defining the terms. Several proposals were made for distinguishing accuracy from appropriateness. It was pointed out by one participant that a statement can be grammatically correct, but still be inappropriate for a given situation. Another participant proposed distinguishing the two terms from a psycholinguistic perspective: Appropriateness is part of the planning stage of speech, whereas accuracy is part of the execution stage. That is, in the planning stage, a speaker chooses a word or grammatical structure based on various sociolinguistic factors, whereas in the execution stage, psychophysical and motor skills are involved, among others. Finally, it was suggested that appropriateness is an interactive term, whereas accuracy is a norm-referenced term. Another participant challenged this statement by claiming that accuracy can be both norm-referenced and sociolinguistically-referenced.

The discussion then turned to the definition of the insiders and the outsiders. It was suggested that the insiders comprise those involved in the adaptation of the government-inspired proficiency test to the academic context, those who train OPI interviewers, those who have gone through the training workshops, and those who are writing the State Proficiency Guidelines. The others would be considered outsiders. Another participant questioned the existence of an ingroup and an outgroup, citing that there are varying degrees of participation, support, and interest in the Guidelines.

The next issue raised dealt with general educational policy. It was claimed that the Proficiency Guidelines are not based on a clearly formulated educational policy for foreign language learning and teaching. The Guidelines were described as an itemized set of independent skills that do not build on one another. It was pointed out that there are no long-term goals for the learning and teaching of foreign languages in this country, whereas in France and Germany, there are explicit philosophical statements specifying what foreign language learners and teachers are supposed to achieve. In Germany, for example, it is stated that through the knowledge of a foreign language, students will learn to critically reflect not only on a foreign language and culture, but also on their own cultural reality. This participant concluded by saying that instead of concentrating our efforts on the revision of the Guidelines, we ought to turn our attentions instead to the development of solid goals and vision for foreign language education.

In reaction to this, one participant commented that although it is true that such philosophical statements do exist in Europe, it is not certain that they have had an impact on what actually happens in classrooms. Another participant suggested that instead of abandoning the Guidelines for the sake of the quest for a general educational vision for foreign language education, we should take a proficiency component and insert it into a much broader framework of what we want foreign language education to be.

The session moved naturally from a discussion of philosophical issues to a discussion of the actual impact that the Guidelines have had on the foreign language classroom, or as Claire Kramsch referred to it in her paper, the "classroom reality". Some of the participants observed that the Proficiency Guidelines and OPI are not used in most classrooms, but that talk of proficiency-oriented instruction has incited teachers to look at what they are doing and to change their view of foreign language teaching to include more communicative approaches. By providing new enthusiasm for communicative language teaching and learning, the Proficiency Guidelines have had a positive effect on the classroom.

The opposite view was also expressed. One of the participants, after having talked with foreign language department chairpersons and teachers around the country, has concluded that the Proficiency Guidelines have instilled not enthusiasm and excitement, but fear, intimidation, and confusion about what proficiency means. These department chairpersons and teachers, along with textbook publishers, believe that proficiency must be important because a national organization, ACTFL, which purports to speak for language teaching in this country, supports it claiming that this is the model to which we should be adhering in evaluating learner progress. They believe that they have to "do" proficiency or they may not receive government funding, school certification, and teacher certification. In addition, it was pointed out that there is even confusion on the inside: Some proponents claim that the proficiency Guidelines ought to serve as a blueprint for curriculum; others claim that they should only serve as an organizing principle for developing curricula because they are not specific enough to do otherwise.

Finally, there was a discussion of the relationship between the Proficiency movement and what Savignon calls communicative language teaching. Communicative language teaching was described by one of the participants as beginning with an assessment of learner needs, followed by the active involvement of learners in the interpretation, expression, and negotiation of meaning in different channels from the beginning of language study. It is a context-specific approach. In addition, it was pointed out that communicative language teaching builds on a firm linguistic base (e.g., the work of Hymes, Halliday, and Firth), and is supported by applied linguistic research by Widdowson and Savignon's research on how learners develop communicative ability. It was emphasized that the important task in this approach is to help teachers and learners to understand the process of the development of communicative language ability.

This approach was then contrasted to the Proficiency movement. It was pointed out that the Proficiency movement is quite different in that it begins with a nationally specified set of guidelines in the form of a test with certified graders. Proponents of Proficiency claim that this represents language acquisition regardless of where, when, and how, and so purports to reflect language learning in a variety of contexts. This participant concluded by saying that Proficiency is not at all reflective of communicative language teaching.

In response to those comments, another participant stated that many accept Proficiency to be almost synonymous with communicative language teaching. Another added that Proficiency and communicative language teaching share similar direction, and the two should attempt to work together.

In conclusion, the discussion in this session covered a wide range of topics related to Proficiency and communicative ability, including a definition of terms for describing communicative ability, the need for educational goals and policy, the impact of the Proficiency Guidelines on the foreign language classroom, and the relationship between the Proficiency movement and communicative language teaching.

THE ROLE OF GRAMMAR IN THE DEVELOPMENT OF COMMUNICATIVE ABILITY

Nina Garrett
University of Illinois

This paper suggests a new model for a principled relationship between (a) grammatical competence and other competences (sociolinguistic, pragmatic, and discourse) and (b) actual communicative language use, and it considers the Proficiency Movement's representation of grammar in these terms. A psycholinguistic model of grammar postulates processing rules (connections between meaning and form) as the primary construct in language use and language acquisition. The paper concludes that the Proficiency Movement's representation of grammar is without psycholinguistic validity and suggests that both L2 acquisition theory and FL pedagogy would benefit from implementation of the concept of grammar as processing rules.

Introduction[1]

The purpose of this paper is to suggest a principled relationship between grammatical competence and the other competences included in the complex which we call communicative ability (usually categorized as sociolinguistic, pragmatic, and discourse competences) and to consider the adequacy of the representation of grammar in the Proficiency Movement in terms of the suggested relationship. (I take the Proficiency Movement to include both the promotion of oral proficiency testing by the Oral Proficiency Interview (OPI) according to the ACTFL/ETS Guidelines and the use of the Guidelines as a framework for organizing teaching practices and syllabus design.)

For some fifteen years now the field of foreign language education has been increasingly concerned both to develop its theory to include an adequate representation of those other competences beyond the grammatical, and to incorporate into its practice techniques for inculcating them in language learners. A heavy emphasis on communicative competence has been -- and continues to be -- crucial in correcting the traditional view that language consists only of a lexicon and a set of rules for combining words into correct sentences. This paper's focus on grammar does not imply any reservation about the importance of the other competences, nor any desire to reassert the centrality of traditional grammar -- quite the contrary. I am concerned, however, that in our zeal to realize the implications of the construct of communicative competence we have allowed the notions of grammar and grammatical competence to slip into an uneasy limbo. Although no one denies that communicative ability must somehow include grammaticality, there is no general agreement in the field as to whether or how the "facts" of grammar as they have always been conceived can be taught so as to transmute into the ability to speak grammatically -- in other words, no principled relationship between grammatical competence and communicative ability has yet been formulated. The postulation of such a relationship requires, I believe, a complete revision of current notions of what grammar is and how it operates in language use. This paper, therefore, addresses the theoretical problems underlying the controversy over the role of grammar not only in the PM but in the field of foreign language education generally.

1. Grammar in testing and teaching for proficiency

The debate over grammar in the PM has been qualitatively different from that over the adequacy of the PM's representation of pragmatic or sociolinguistic or discourse competence. Discussions of the latter three have concentrated on three questions: (a) whether the Guidelines are based on a sufficiently sophisticated concept of the particular competence, (b) whether the OPI can elicit a language sample adequate to give ratable evidence of that competence, and (c) whether a syllabus based on the Guidelines can inculcate it. The argument, in short, has been whether the PM represents an adequate understanding of the concepts included in communicative competence and steers closely enough towards them.

As regards grammatical accuracy, in contrast, the debate has focused on the questions (a) how much it should be weighted in the testing of global proficiency, and (b) whether there are either theoretical or empirical grounds for allotting much (if any) classroom time to teaching grammar. It is taken for granted that the concept of grammatical competence is understood; the argument has been whether the PM steers far enough away from it.

2. The issue of grammar in the field at large

In fact this debate is not new to the field. The question of whether or not grammar should be explicitly presented has plagued the field for centuries (c.f. Kelly, 1969) and has been a central issue in debates about communicative competence since that concept became an organizing principle for foreign language education some 15 years ago. It is not really an issue which constitutes a dividing line between those who adhere to the notion of communicative competence as an organizing principle and those who prefer the notion of proficiency. Both sides pay lip service to the claim that grammatical competence is an essential part of communicative ability,[2] but there is no generally accepted theoretical or methodological stance on how it is to be achieved. The arguments are usually divided along the lines of (a) "there's not much reason to believe that teaching grammar does any good, and some reason to believe that it could be harmful" as opposed to (b) "there's not much reason to believe that not teaching grammar works any better, and some reason to believe that that could be harmful," and each side attacks the research cited by the other. The PM includes "accuracy" as one of the three major components of communicative ability[3] and is generally seen as promoting attention to grammar, but there seems to be a wide range of opinion in both camps as to the desirability of teaching grammar explicitly.

Underlying the debate, however, is a striking similarity of fundamental assumptions: both those who believe that teaching grammar is necessary for classroom second language acquisition and those who believe that it is not only not necessary but generally not even advisable are generally in agreement on several fundamental points:

- that the nature and function of grammatical competence is established and understood;
- that to communicate grammatically one must somehow acquire the rules of competence, but that one can and should do so without too much overt attention;
- that the nature of communicative competence is generally understood, though the particulars are still being developed;
- that communicative competence is more important, more interesting, and requires far more attention than grammatical competence.

This paper suggests that none of these beliefs can be taken for granted, and that they form an inappropriate set of preconceptions about what grammar is and about the role it plays in classroom second language acquisition.

3. The Basis for the role of grammar in the Guidelines

The Guidelines as rating level descriptions were originally developed on the basis of experiential data, reflecting the job-related language requirements of the population being tested, rather than on the basis of adult classroom second language acquisition theory or theoretically motivated empirical research. The specification of the sequence in which certain structures are expected to be produced is dependent on the ordering of the functions, and expectations for the development of accuracy are based on the production of large numbers of learners. This needs no defense: experiential data constitute an altogether reasonable starting point for developing such a test. As long as the OPI is being used as a test of certain aspects of what we teach, it need only have reliability and face validity; as long as those who use it agree that the test results correlate to their satisfaction with learners' ability to function in the FL under specified conditions and for specified purposes, it might be argued that no independent theoretical basis is required for the specifications of accuracy.

Again, if the Guidelines are to be used to guide syllabus design, the crucial question is the justification claimed for so doing. We are certainly at liberty to decide that we want our students to be able to carry out certain functions in certain contexts, and to specify the sequence in which we want to present these functions and contexts explicitly in class. If we agree on the specifications for a syllabus, it too needs no theoretical basis. If we want students to be able to translate, we design a syllabus to inculcate that ability; if we want them to carry out certain communicative functions, we design a functional syllabus. Deciding on a syllabus is a matter of pedagogical policy.

But it seems to me that current developments in the PM are based on a confusion of pedagogical policy with theory. If the Guidelines are used to motivate syllabus design and teaching practice not only because they furnish a functional description of a specified pedagogical goal but also because they are believed to describe how adults in the classroom learn a foreign language, then a theoretical claim is implied which requires independent support. Without a theoretical basis, recommendations for syllabus design and teaching practice based on experiential data are inevitably circular: they imply that because most learners in the past seem to have learned this way, this must be the natural way for them to learn and therefore this is the way they should be taught. This circularity becomes genuinely dangerous when it motivates suggestions to the effect that since the Guidelines indicate that most learners don't seem to be able to use grammar very well until later levels, maybe there's no point teaching grammar until they get to those levels.

It might be argued that most proponents of the PM do not in fact make any theoretical claims, and it is true that the field of foreign language education as a whole tends to pride itself on being atheoretical, on its eclecticism, on "caring about what works." But the apparent unresolvability of the arguments about the value of explicit grammar indicates that we often do not know what works, and we cannot undertake research on what works as regards grammar without developing a coherent and testable theory of its role in the development of communicative ability. Partly because grammar has become a fighting word, and partly because we have always assumed that we knew what it was, no one has wanted to undertake to do so, but the momentum gathered by the PM makes it absolutely necessary. Without a theory of how grammar works the PM has no substantive basis for claiming that the OPI elicits or assesses grammatical competence in any useful sense, still less a basis for claiming that the Guidelines provide any guide for whether, or how, to teach it.

Therefore the problems connected with assessing the role of grammar in the PM are not so different from the problems attending the other components. The debate about the role of grammar should not be how we can downplay it, but should be of the same form as the other debates -- whether the concept of grammar which informs the Guidelines is adequate to serve as one of the structural members of a framework for syllabus design or to assure that the OPI can elicit data which will allow us to assess, in any interesting way, the grammatical competence of our learners. At the moment the answer must be negative, but that is the field's failure as much as the PM's.

4. Problems with current research

It is beyond the scope of this paper to review the three related fields of research to which we have looked for information on how grammar works (foreign language education, natural second language acquisition, and first language acquisition) but I shall suggest reasons why the research done in each has been of little use to us in establishing a useful understanding of the role of grammar.

First, in foreign language education, research on the topic of grammar has been inconclusive because both sides in the debate over the teaching of grammar have reasonable grounds for challenging the other's position. The studies often used as a basis for insisting that "grammar is necessary" (reports on fossilization, the terminal 2, and native speaker reactions to learners' ungrammatical utterances) have been found unsatisfactory on a number of grounds by VanPatten (1986). On the other hand, however, the success of teaching methods which eschew grammar is equally questionable on both empirical and theoretical grounds. As to the practice, at the early levels of learning any method can be highly successful in the hands of experienced and enthusiastic teachers; no substantive claim about the course of long-term acquisition can be made on that basis. As to the theory, the arguments against teaching grammar which are commonly cited in contemporary literature are derived from Krashen's work, which has been persuasively criticized by Spolsky (1985) among others.

Second, natural second language acquisition research can make only a limited contribution to our field of research. By far the largest proportion of it has been carried out on the learning of English, and the inappropriateness of extrapolating from that to the learning of other languages has not been adequately realized. More important, there is no evidence that I know of to indicate that the adult natural second language learners whose data are typically used as a basis for generalizing about the course of second language acquisition actually do go on to achieve levels of second language communicative ability which we would find acceptable for classroom learners. For all we know, the course of wholly untutored adult second language acquisition may typically lead to a dead end: the early stages of such acquisition may be characterized by cognitive processing altogether different from that which will eventually lead to the kind of communicative ability we project as a pedagogical goal.

Third, research on first language acquisition is of limited use to us on its own terms for a number of well-known reasons. Too many of the variables are different -- age, cognitive maturity, motivation, classroom constraints, etc. And having a first language really does make a difference -- the pendulum has now swung back far enough from the anti-transfer position of the 1970's to allow that to be said in public. Nonetheless research on first language acquisition is of crucial importance to us -- not that knowing how children learn their first language tells us how classroom adults learn a second language, but rather that the questions asked by first language acquisition research suggest important hypotheses which should be explored in our domain.

5. Theoretical assumptions underlying the research

The major reason why research in foreign language education and natural second language acquisition has not been helpful to us as regards grammar is that it has been based on the premise -- which this paper challenges -- that to explore the role of grammar in L2 acquisition (classroom or natural) means to study how learners "use" those rules of morphology and syntax which constitute the traditional linguistic description of a language -- how learners acquire those rules, whether they need to be taught them explicitly in order to acquire them (or whether explicit learning gets in the way of automatic use), what makes them available underlyingly for use, under what circumstances learners can access them, and so on.

Those morphological and syntactic rules are generally said to constitute Chomsky's "competence," but strictly speaking competence does not comprise the rules but only the intuitions of native speakers about how the grammatical structures of their language are related to each other in a system; it is from these intuitions that the abstract rules or generalizations are derived. When we talk about the relationship between the abstractions contained in competence and the actual behavior involved in speaking and comprehending language (performance) we use phrases like "underlying knowledge," or "using the rules of competence" as a "basis for performance," and these phrases have become "metaphors we live by" (Lakoff & Johnson, 1980), metaphors which themselves structure our perception of what we are investigating. As a result we have misguidedly attempted to specify the role of grammar in the acquisition of communicative ability in terms of working out how learners acquire grammatical competence in a foreign language as the basis for their performance in a foreign language. The concept of communicative competence has been added to represent the notion that there is more to speaking a language than implementing grammatical knowledge; in another metaphor we continue to live by we often characterize the rules of communicative competence as "telling us how to use" the rules of grammatical competence, as mediating between it and performance.[4] These metaphors have thus provided a rough general framework for models of language acquisition. (See Figure 1.)

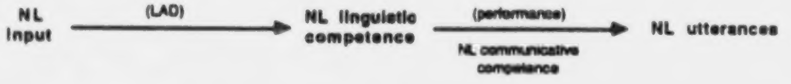


Fig. 1: A model for a model of native language acquisition.

These metaphors have made competence the central construct in second language acquisition theory: competence is thought of as that-which-is-acquired because it is thought of as that-which-is-used, as the basis for performance -- and we have explicitly insisted, since the demise of the "grammar-translation method," that performance is the goal; we want our students not merely to know about the language but to be able to communicate with it.

The centrality of competence in the model is the crux of our problem. It is not useful for us to investigate the role of grammar by studying how learners acquire, access, or use the traditional linguistic rules of a language, because the linguistic rules are not the basis on which anyone -- native speaker or learner -- actually produces or comprehends an utterance. Obviously, unarguably, the goal of language learning is that the learner should come to be able to produce and comprehend utterances which fit the linguistic description of the language being learned; it cannot be suggested that the linguistic notion of a grammar has no role to play in accounting for language use or language acquisition. I do suggest that allowing competence to occupy a central position has confused our attempts to develop a coherent theory of the role of grammar in classroom second language acquisition.

6. The relationship between meaning and competence

In producing or comprehending an utterance a speaker-hearer doesn't start with competence. The basis for language behavior is not a set of intuitions about how a language system as a whole works: the basis for performance is the desire to express some meaning. Meaning is encoded into utterances through a complex set of processing rules, psycholinguistic rules which operate in real time in speakers' minds to combine lexical items and a variety of formal features (word order, inflections, etc.)[5] The linguistic facts are psycholinguistically relevant only as they are present in the mind of speakers, as they are used in processing. But how are grammaticality intuitions used, if they are not the basis or starting-point for processing? I suggest that they constitute instead the end-point, that competence can best be understood as a template, a sort of Platonic ideal of the particular language, against which any utterance has to be mentally matched. Figure 2 presents a model of psycholinguistic processing.

The conventional model postulates competence as the basis for performance. Here, in contrast, we begin with meaning, but not just with semantic content; pragmatic,

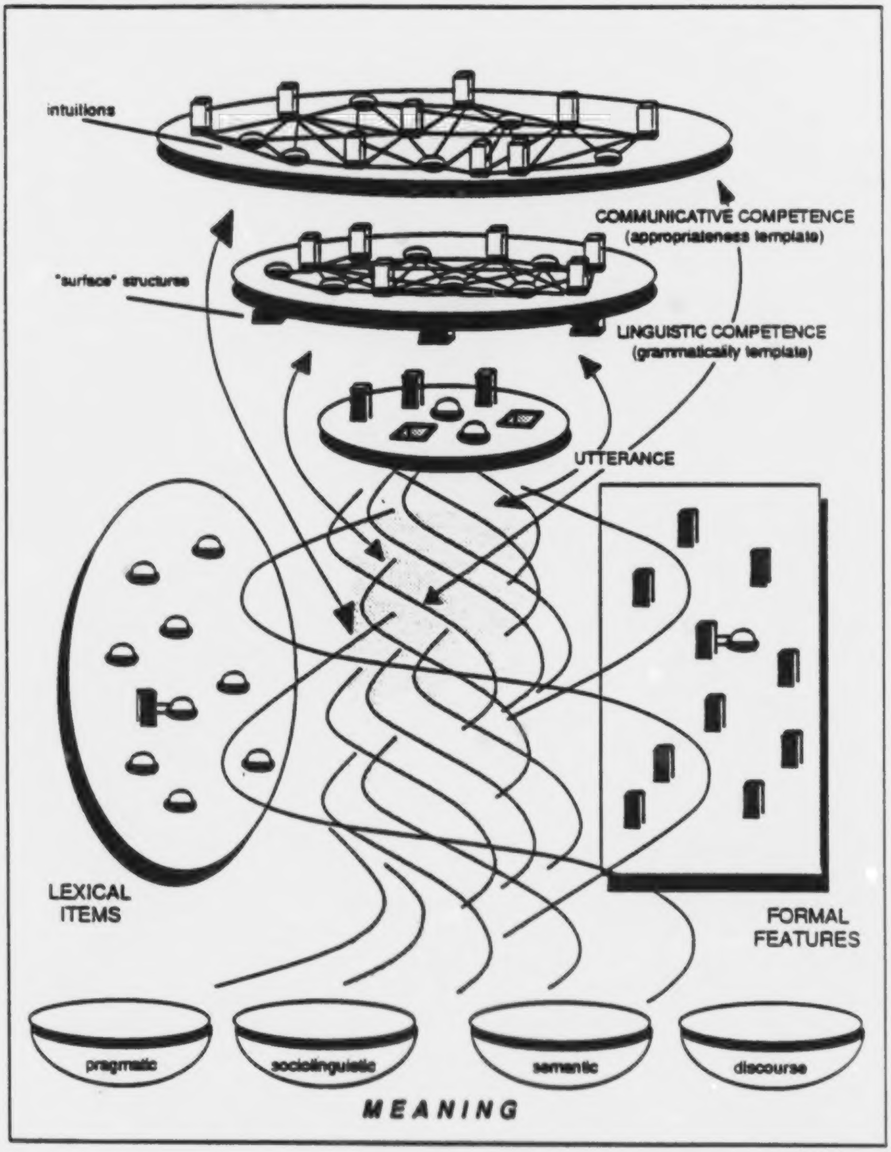

intuitions
COMMUNICATIVE COMPETENCE
(appropriateness template)
"surface" structures
LINGUISTIC COMPETENCE
(grammaticality template)
UTTERANCE
LEXICAL ITEMS
FORMAL FEATURES
pragmatic
sociolinguistic
semantic
discourse
MEANING

Fig. 2: A model of native language processing.

sociolinguistic, and discourse meaning are integral to every adult utterance. The swirling lines (the processing rules) working upward from meaning, loop through the speaker's repertoires of lexical items (some of which require the attachment of specific grammatical form) and formal features (some of which require the attachment of specific lexical items), combining them in an utterance whose configuration must fit (if the utterance is grammatical) into the template of grammatical competence, the native speaker's intuitions about the linguistic system. As for the relationship of communicative competence to grammatical competence, this processing model represents the claim that communicative competence includes grammatical competence (an utterance can be grammatical without being appropriate, but it cannot be appropriate without being grammatical) but goes beyond it. This use of the term competence is exactly analogous to the original: communicative competence refers to another kind of knowledge, another set of intuitions, above those which confirm grammaticality, about what is appropriate, and it thus functions as another template, a set of supra-grammatical constraints on language behavior. If the utterance is grammatical without being appropriate, in the terms of the model we can say that it fits the linguistic template but does not fit the communicative template. The arrows which flicker upwards towards both templates during the processing of an utterance indicate the almost entirely unconscious mental matching, the check "whether it's going to come out right" which can be recognized even in fluent native speakers by hesitations, slips of the tongue which produce ungrammatical utterances, and self-correction -- all of which apply to mental checking for various kinds of appropriateness as well as for grammatical accuracy.

This processing model is a thus visual metaphor for the claim that both grammatical and communicative competence have psycholinguistic status, but not as the basis for performance. The linguistic rules of a language are derived from native speakers' intuitions about what is grammatical; the discourse, sociolinguistic, and pragmatic rules of a language are derived from native speakers' intuitions about what is appropriate. All of these rules function as a body of knowledge against which the result of the processing -- the production or comprehension -- can be checked, not as a starting point for, nor as the central process in the actual production or comprehension of utterances.[6]

7. "Teaching Grammar" re-examined from the processing perspective

Once we see processing rather than competence as the central construct it is easy to understand why the explicit teaching of target language linguistic rules is not particularly successful in assisting learners to acquire the ability to communicate grammatically. Linguistic rules are not per se processing rules, so explaining the linguistic template (telling learners what rules their finished utterance must conform to) will not of itself tell them how to construct those swirling lines, the processing rules connecting meaning to its formal realization.

However, because the two sets of rules are related, language learners can often work from one set to the other. Some classroom learners can use even "old-fashioned" grammar instruction to develop communicative ability, which means that they can extrapolate processing rules from linguistic rules, i.e., extrapolate meaning-form connections from discussions about the abstract relationships between formal features. When, in contrast, the presentation of linguistic rules is sharply reduced or avoided, the "good" learners are those who can extrapolate those meaning-form connections from input instead. (Since input is not always directly or unambiguously meaningful the learner is in effect being asked to hypothesize a connection between form and meaning without either being specified.) Learners differ in their preference for inductive vs. deductive learning, and it may be that those who prefer the former will acquire communicative ability better without explicit grammar than those who prefer deductive strategies. The fact that some research on the efficacy of explicit grammar is inconclusive may be due to a roughly equal representation of learners of both types in a study; similarly a strong finding against explicit grammar might result from an accidentally high proportion of inductive learners in the subject population. (I do not know of studies along these lines in which learner preference for inductive vs. deductive learning was controlled for.) The point is that the processing perspective allows us to ask an entirely different question: would it help students if the processing rules were explicitly explained so that they would not be forced either to induce or to deduce them? This question will be taken up later.

In the short run, what works may be simple surface structure collocations and heuristics which are virtually empty of meaning. But in the long run any syllabus or method that aims to enable learners to do something with the language rather than only to know something about it has to make some claim (even if only implicitly) about all the components of language performance -- the meaning, the formal features, and the processing which connects them. Different methods or approaches will, of course, place their primary focus on different components. In terms of the processing model, the grammar-translation method focused on the rules in the linguistic template (though of course it did not pretend to teach communicative ability). Communicative approaches focus on sociolinguistic, pragmatic, and discourse meaning, avoiding excessive attention to the linguistic template -- though different versions of the communicative approach will differ as to what is considered excessive. Some variants of a communicative approach regard sociolinguistic meaning as primary, while a functional approach often places special emphasis on the pragmatic. A notional approach in principle stresses semantics, but the "notional-functional" syllabus tends in practice to be mostly functional, if not outright situational. The processing model shows all four kinds of meaning at the same level; one or another kind of meaning may be primary in any given utterance, but every utterance represents the interaction of all four.

One of the major problems inherent in the development of theory is that it abstracts from the richness and complexity of empirical reality and gives rise to the polarization of positions. (Such polarization needs to be recognized and qualified, lest it lead to counterproductive controversy, but it can also lead to a sharpened awareness of underlying assumptions and theoretical contrasts.) The actual classroom practice of most teachers is based less on any theoretical position than on a pragmatic and adaptable set of hunches (or preconceptions based on their own language learning experience) about what seems to "work," even if they are not trained or inclined to think about the form-meaning connections actually involved in basic grammatical notions. These capsule descriptions of syllabus philosophies thus leave out of account the practice of teachers who reject labels and party lines and work out their own compromise position on grammar. Some, for example, feel the need for a grammatically structured syllabus but seldom if ever explain a grammatical structure explicitly for its own sake, preferring to focus students' attention on the communicative import of a structure by highlighting its use in communicative contexts. Others reject a grammatically structured syllabus in favor of a functional one, but may provide explicit explanation and drill of whatever grammatical structure arises in connection with a given function, perhaps providing it entirely outside of classroom time via workbook or computer. Many other compromise positions can be imagined and are undoubtedly practiced by highly experienced and successful teachers, and their intuitions about what works cannot be chopped or stretched to fit any Procrustean bed of pedagogical theory about explicit grammar.

In arguing for the importance of a psycholinguistic rather than a linguistic theory of language processing I may myself have fallen prey to that tendency to polarize theoretical contrasts. I should therefore reiterate the close relationship which exists in principle between the linguistic and the psycholinguistic and is clearly postulated by the visual metaphor. In practice, however, conventional textbook explanations of linguistic rules are usually couched in terms either of grammatical terminology (under the assumption that terms like "direct object" or "tense" are meaningful to the learner), or of correspondences to the meaning allegedly conveyed by a given form in English, or of correspondences with another formal feature (indicating for example that the passive means the same thing as the active) and these explanations cannot in fact be assumed to convey any concept whatever to the learner. Furthermore, even when a form is explained immediately in terms of some bit of meaning, the value of the explanation is frequently nullified by explicit or implicit indications of surface structure clues by which the learner can guess what form is called for without having to analyze meaning. As a result learners develop, as interlanguage rules, heuristics for using TL forms to convey an element of meaning quite different from that which the TL actually marks with them, although the differences may be so subtle that the learner utterances containing those forms are correct much of the time.[7]

Obviously there are teachers who have developed a variety of ways of presenting linguistic rules so as to make them maximally translatable by learners into processing rules, and obviously there are learners who have little difficulty working from linguistic rules to develop communication strategies and appropriate early-interlanguage rules. We have always known that good language learners can learn well via any approach, and we have always known that good language teachers can teach well no matter what theory or methodology they are trained with. But that does not guarantee that any teacher can with a given amount of experience develop the requisite insights and techniques. I suggest that the concept of processing provides us with a principled basis for understanding what we already know intuitively about good learning and good teaching, from which we can generalize to help less good teachers and learners.

What would be the implications for our teaching practice of a more fully developed theory of grammar-as-processing-rule? The most immediate is the importance of recognizing the integrity of processing, the integrity of communicative language ability, which has been repeatedly stressed by other papers and comments throughout this symposium. The separation of form and meaning is not only impossible in terms of all that we know about the nature of language knowledge and language use but also counter-productive in terms of our philosophical goal of helping students express their own meaning in the appropriate forms of the target language. The separation of grammatical and communicative competence is equally impossible: meaning of any kind -- semantic, relational, propositional, sociolinguistic, pragmatic, discourse -- can only be conveyed through appropriate use of grammatical form, at least beyond the single-word level. Nor can the processing of the lexicon be separated, except in superficial terms, from the processing of grammatical form; there is no fundamental difference between learning how lexical items convey bits of meaning (all those various kinds of meaning) and learning how bits of form do so.

But this discussion of the importance of conceiving of grammar in terms of the processing model should not be taken to imply that processing rules can be handed to students on a silver platter -- that explaining grammar in terms of the form-meaning connections will instantaneously enable learners to speak grammatically. Processing cannot be directly taught so as to guarantee production that is either grammatically accurate or communicatively appropriate, any more than grammaticality can be taught just by explaining linguistic rules. Nonetheless, the concept of processing itself is surely crucial as a framework within which learners can organize the language data they get from communicative input or any other source. Beyond presenting that, one could suggest three steps in "teaching" processing rules which are realized in grammatical form: a) making students aware of a particular bit of meaning (of whatever kind) which must be conveyed grammatically in the TL; b) setting up the form-meaning connection at whatever level of refinement is appropriate for the language-learning level; c) providing ample opportunity to practice the processing rule.

Notice that "making aware" and "setting up" and "providing opportunity" are all deliberately vague as to the degree of explicitness. In each language, for each form-meaning connection, decisions about what degree of explicitness is required should be motivated by research exploring the adequacy of the interlanguage rules students are likely to develop under various approaches. Traditional assumptions about the transparency of form-meaning connections and students' ability to induce or create them are often wrong. Because we have barely begun to study learners' interlanguages from this point of view we do not yet know which early crude interlanguage rules are likely to be useful and inherently refinable as opposed to those which are likely to be dead-end and a permanent hindrance to the construction of a TL-like processing rule.[8]

The global implication for foreign language education of understanding grammar in terms of processing is that we must change our efforts to focus on learning rather than on teaching -- and we must address the individual learner's needs. There is no best method or best approach for all classroom foreign language learners. Differences in individual learners' affective characteristics have been recognized for years now, and differences both in cognitive style and learning style are receiving increased attention. The realization that individual learners' creation of interlanguage processing rules is also highly idiosyncratic makes inescapable the conclusion that the only viable approach to grammar is (1) to help the students themselves to think about language acquisition and language use in processing terms so that they can understand their own learning needs and predilections, and (2) to maximize the individualization of learning options for them.[9]

8. The problem of sequencing

Even with a theoretical framework motivating the recognition of structures as coding elements for the processing of meaning, we still need principles for deciding on a sequence of presentation. Neither child acquisition orders or those of natural learners can be taken a priori to be appropriate for the classroom. Is there any theoretical criterion for structural sequencing?

One approach is suggested by the work of Talmy Givón, who insists that language can only be understood from a functional perspective, a claim which should recommend his work to us. He posits two extreme poles on the continuum of native speakers' communicative language use, the pragmatic mode and the syntactic mode, and explicates a number of different relationships between them (Givón, 1979). The pragmatic mode is characterized as having topic-comment structure, loose conjunction, slow rate of delivery, pragmatically governed word order, and little or no use of grammatical morphology, in contrast to the syntactic mode's subject-predicate structure, tight subordination, fast rate of delivery, word order governed by semantic case function, and elaborate grammatical morphology. Givón argues that the contrast between the pragmatic and the syntactic obtains not only diachronically in the development of a given language (and to a certain extent his arguments on this point fit in with Slobin's explication of "Language Change in Childhood and History," 1977) but also between pidgins and creoles, between adult and child language, and between informal/unplanned speech and formal/planned language. He explicitly mentions the relevance of this distinction to the adult second language learner: "... under a pressing need to communicate in a foreign language, the adult does, in fact, revert almost all the way to the extreme pragmatic register on the scale."

To a remarkable extent Givón's arguments overlap with those made by Corder in his paper "'Simple Codes' and the source of the Second Language Learner's Initial Heuristic Hypothesis" (1977). The simple code which Corder postulates is clearly much the same as Givón's "pragmatic mode," and indeed Givón's terminology solves the problem, which Corder addresses at some length, of the undesirable connotations of "simple" and "simplified." Many of Valdman's (1980) recommendations for a modified structural syllabus are based on similar analyses of relationship between pragmatically motivated paratactic constructions and the syntacticized ones mandated by traditional textbook presentations. The similarity of these claims and their attempts to account for quite different problems and data (although they use different terminologies in their discussion) suggests that this concept could form a significant connection between linguistics, second language acquisition and foreign language education, and could provide us with testable hypotheses about the order in which syntactic and morphological features appear in learner language.

Givón's theoretical framework is particularly suitable for our current priorities in that he insists on the primacy of discourse and the need to consider grammatical rules as an automatic processing strategy needed to effect communication under certain discourse circumstances. (This emphasis on discourse constraints as organizing grammar is also an important feature of Karmiloff-Smith (1981) whose subjects produced quite different grammatical features in connected discourse as contrasted with their single sentence utterances.) An analysis of the "progressively-more-syntacticized registers" posited by Givón could provide us with a more principled set of guidelines for what we might expect in second language learners, especially at intermediate and advanced levels.

However, the role of discourse in organizing the acquisition of grammatical processing should not be over-emphasized, for several reasons. First, discourse constraints may dictate which of two possible grammatical structures is appropriate, but they do not normally allow the choice of a structure which is ungrammatical at sentence or clause level. Second, too heavy a focus on discourse may prevent us from recognizing learner problems with basic grammatical relations and other primary building blocks of simple clauses. Our textbooks routinely assume that these pose no particular problem, presumably because it is taken for granted that even learners who have not been exposed to advanced grammatical terminology are familiar with the concepts represented by terms like subject, direct object, and indirect object. Garrett's research (1982) indicates that this assumption is groundless, since even "good" language learners often have conceptually mistaken interlanguage rules for these. Traditional explanations of grammatical relations (to whatever extent they are given at all) usually present a confused conflation of semantic and syntactic properties, failing to indicate (either explicitly or through coherent presentation of examples) that there is a difference between semantic and syntactic roles, and that languages vary considerably as to the assignment of semantic roles to syntactic case. (English, for example, allows both beneficiary and experiencer to be represented as grammatical subject, but most of the commonly taught languages allow neither.) Even the processing of such basic notions as time and space is by no means transparent, even when both the notions and the encoding are similar in the native and target languages.

9. Summary

The cumulative import of the theory and the research cited here is to suggest that our theoretical and practical problems with the role of grammar in the development of communicative ability might be effectively addressed if we adopted a perspective establishing that:

- processing or performance, rather than competence or knowledge, is the central construct in acquisition;
- interlanguage is a set of processing rules rather than knowledge rules
- the processing of grammatical form is inseparable from that of lexical items;
- grammar is not separate from meaning and should not be seen as a formal entity to be taught or assessed separately from the communication of semantic, sociolinguistic, discourse, and pragmatic meaning;
- linguistic and communicative competence are both knowledge structures, not abilities.

This discussion of the processing perspective on grammar should suggest my reasons for dissatisfaction with the way grammar is represented in the Proficiency Movement. As I said earlier, its inadequately principled basis for assessing grammatical competence within communicative competence is representative of the inadequacy of the field as a whole in this regard. There is good reason for the absence of grammatical specifications for the realization of communicative functions in the generic guidelines, but the necessity for language-specific ratings descriptions poses a real dilemma. Not to include grammar specifications is to ignore the form end of the form-meaning connection, but to include them is to base ratings on the assessment of product rather than process, and moreover tempts untrained teachers to use the OPI as a discrete-point elicitation task. A more serious problem is that the levels are inherently sequential and implicational, while the acquisition of processing rules is not. There is no way around the achievement-test quality of the grammatical expectations at the lower proficiency levels, and no way around the circularity of the level descriptions vis-à-vis the learner's supposedly syllabus-independent ability at any level. These points have been made before in relation to the assessment of other competences.

To address the problem of grammar more specifically in terms of the processing model proposed in this paper, we must consider what we can actually assume about a learner's ability to create functionally appropriate and appropriately accurate form-meaning connections on the basis of a TL interview. As to the meanings involved, since we are dealing with adult learners we cannot assume any predictable progression in the complexity of the meanings they will want to and try to convey; they are cognitively quite mature if not yet fully skilled (even in the NL) at hypothesizing about abstractions. That is, we cannot assume that a second-year student is attempting only to convey notions for which the grammatical form is taught in the second year. As to the formal features, I have argued that the appearance or non-appearance of particular forms in their obligatory contexts (linguistic or communicative) cannot be assumed to give us useful information about what the learner actually means by the form. As several papers presented at this Symposium have pointed out in other contexts, there is no psycholinguistic basis for the rating descriptions.

I must reluctantly conclude that the OPI is inadequate of itself to test oral proficiency and that the use of the Guidelines to shape syllabus design and teaching practice is unjustified. I do believe that the OPI is a usable oral elicitation device and that the data it yields could be used together with other data to contribute to a teacher's sense of a student's ability to perform in one kind of discourse situation. There is no question that the training sessions and workshops familiarizing teachers with the notion of oral proficiency and the multiplicity of factors which go into it have given teachers a more realistic understanding of the goal of communicative language ability and have given them techniques -- to be adapted to their own students' situation -- for assessing some aspects of what they are trying to teach. Beyond that, however, the uses to which the test and the guidelines are widely being put seem to me to be unwarranted.

10. Research

It is part of the protocol to end a paper like this one with the call for more research, but it is pointless and potentially divisive to call for research without addressing the questions of "what kind of research is most urgently needed? how can it be carried out, and by whom? what can we expect to do with the results?" This paper concludes, therefore, by positing some desirable features of research and theory development as regards the role of grammar.

- Above all we need to understand second language acquisition in our own terms, as it takes place in the adult foreign language classroom, rather than extrapolating from research on language acquisition under altogether different circumstances. Not only is this essential for our pedagogy, but it is crucial for the integrity and intellectual status of the field that we demonstrate that acquisition in our domain has as much theoretical interest as acquisition elsewhere. (It goes without saying that our theory must be consonant with the language acquisition theory developed for other domains.)
- We must clearly understand the distinction between "research in second language acquisition in the classroom" and "research in foreign language education" and the priority of the former. That is, we must have some idea of what learners are actually doing with language before we can study the effect of manipulating pedagogical variables. There is an enormous amount to be learned by holding pedagogical variables constant while we investigate the processing of individual learners with different cognitive and affective characteristics, in different languages, at different levels, in different tasks, etc.
- Nonetheless, it is by definition impossible to study classroom

acquisition free of pedagogical manipulation. It is misguided, therefore, to study learners' processing at any point in time, or its development over time, with the aim of "showing what learners naturally do," or to conclude from any such data that because some particular development can be commonly seen it is natural or inevitable and cannot or should not be pedagogically influenced. Most of the arguments against explicit grammar rest on the illogical conclusion that because the "application" of explicit traditional linguistically based grammar has a limited effect on learners' communicative ability, we shouldn't try to teach any kind of grammar or attempt any kind of error correction. The suggestion that explanation or error correction based instead on the concept of processing will have more effect is a testable hypothesis. Perhaps there are certain stages of processing (or certain structures or certain kinds of meaning) which will turn out to be resistant to pedagogical manipulation -- if so it would be extremely interesting to know what they are and why -- but that certainly cannot be assumed on the basis of research to date.

- Research must focus on the individual learner or on particular characteristics by which data from individuals can validly be grouped as well as on data aggregated across groups. There are too many individual characteristics, such as cognitive styles, affective traits, and language learning backgrounds, whose influence on processing is as yet unknown. Even within groups of apparently highly homogenous learners interlanguage processing rules differ dramatically (c.f. Garrett, 1982). Generalizations about language acquisition by groups must be complemented and informed by more information on individuals.
- Once we have the theoretical framework and methodological sophistication to investigate processing rules we can study their differential development under a variety of pedagogical treatment. We need to study three approaches, all in classrooms organized from a communicative perspective (1) where "grammar" is not explicitly taught in any way, (2) where grammar is taught as it is now in communicative classrooms whose teachers do believe a little is worthwhile, and (3) where the grammar presentations are explicitly organized from the processing perspective.
- It is imperative that this kind of research program be a joint effort, with those whose primary concern is classroom practice working closely with those whose primary concern is the development of theory and research. Most of the participants in this Symposium are equally committed to both endeavors, but it is not necessary that every teacher be a theorist or vice versa. It is necessary that we work together. The theorist needs the teacher, because, unlike linguistic predictions, psycholinguistic claims require empirical verification; the teacher needs the theorist, because non-theoretical notions of "what works," no matter how plausible or persuasive, quickly become bandwagons and as quickly become passé.

There are many reasons, too well-known to require listing here, why the undertaking of classroom second language acquisition research has been so difficult for all sectors of the field. However, the two most important -- the "junk field" perception that makes pedagogical research unpromotable in many literature departments, and the extraordinarily heavy time commitment demanded of language teachers at every level -- are already beginning to change.

As to the first, the theoretical perspective suggested here has the potential for changing the academic perception of the field, and foreign language departments are moreover being pushed to recognize the validity of acquisition research by the fact that they need more language faculty than literature faculty. If we can work out principled relationships between a theory of classroom second language acquisition and the theories which underlie work in linguistics and psycholinguistics, then we will be part of a discipline which can lay claim to intellectual rigor, and the research which builds that discipline should be promotable by any reasonable standard a research university cares to apply. This emphasis on basic research does not imply denigration of methodological research, nor are the two in competition. Both must go on -- independently, in parallel, and hand in hand.

As to the second barrier to research, time constraints, the new factor is the computer. The potential of computer-assisted instruction in foreign language education is well-recognized though as yet poorly realized. (I have discussed elsewhere (Garrett, 1987) the possibilities for realizing the concept of processing grammar in computer-assisted instruction.) The enormous potential for computer-assisted basic research, however, is as yet hardly dreamt of. Here I do not mean research designed to evaluate the efficacy of computer-assisted instruction, though studies focusing on particular aspects of that question are certainly needed, but rather second language processing research which uses the computer to elicit, collect and analyze learner language data. The technology has for some years already been capable of supporting processing research on complex written data (c.f. Garrett, 1982) and computer-controlled audio and video technologies now allow in addition aural and visual manipulation of both stimulus and response. All it takes now is for teachers to recognize the potential and learn how to use it. It is the conceptual design of the research, not the implementation, which is difficult.

The research agenda outlined here is considerably broader than an investigation of the particular topic assigned to me, the role of grammar in the development of communicative ability, but I hope that my colleagues who are discussing other aspects of communicative ability will not think that I am encroaching on their territory if I suggest that the organizing principle of processing allows us to see that grammar in its widest sense includes all the ways in which we put words together to communicate meaning. In my opinion the most serious challenge facing the Proficiency Movement is the need -- made explicit in virtually all the presentations of this Symposium -- to build into the effort to improve teaching and testing an ongoing program of research and theory development dedicated to a holistic concept of language acquisition, which will, I hope, finally integrate a principled concept of the role of grammar.

Notes

[1] I am indebted to all those who spoke at the Symposium for the clarification of issues, sharpening of focus, questions and criticisms which contributed to the final version of this paper. In addition I am indebted to Gary Cziko, John Lett, Alice Omaggio, and Bill VanPatten for their comments on the earlier written version prepared for pre-Symposium circulation to participants. Any inadequacies or heresies remaining are my responsibility.

[2] Canale and Swain (1980) give a careful overview of various theories of communicative competence which contribute to our present understanding of it, and they distinguish between those theorists who consider communicative competence to include grammatical competence along with sociolinguistic, pragmatic, and discourse competences and those who see the communicative as separate from, complementary to, the grammatical. As will become clear later, I adhere to the former.

[3] In the discussion of several papers presented earlier in this Symposium the meaning attached by the PM to the term "accuracy" has been the focus of considerable debate. Omaggio (personal communication) and Byrnes (1987) both insist that it is intended to include accuracy of vocabulary and pronunciation as well as of grammar, and appropriateness in terms of sociolinguistic and discourse competence as well. However, in foreign language education generally as well as in linguistics "accuracy" usually refers to grammar and "appropriateness" refers to sociolinguistic, pragmatic, and discourse factors, and the terms will be thus used in this paper.

[4] This metaphor itself implies the second view of communicative competence discussed in Note 1 above, the notion that communicative rules are separate from and added onto grammatical ones; this view is explicitly rejected by Hymes (personal communication) among others, myself included.

[5] In comprehension the processing rules operate to decode utterances into meaning, though probably not in a crudely reversed fashion. In the interests of simplicity "processing" here will refer to production, but of course comprehension is equally important.

[6] This model is explicated in more detail in Garrett, forthcoming.

[7] See Garrett (1982, 1986) for a fuller development of this argument.

[8] The practice referred to in the third step has to be at several levels. In paradigmatic practice learners do not need to make decisions as to which kind of form is required but only as to which piece of the requisite paradigm is required by the surface structure; this practice focuses on the formal end of the connection. The meaning end is strengthened by practice in which choices between forms must be made on the basis of meaning, for example whether the aspectual connotations require the perfect or the imperfect, or which of two objects represents the beneficiary. Finally, in open-ended communicative discourse, speakers must make both syntagmatic and paradigmatic choices on the basis of semantic, sociolinguistic, pragmatic, and discourse meanings simultaneously.

Pedagogical common sense as well as theoretical plausibility suggest that these levels of practice should at first be controlled for the student. In conventional grammar teaching there is seldom any intermediate step between discrete-point paradigm drills and open-ended discourse, and therefore the student cannot profit from error correction: the mere pointing-out of the inaccuracy or inappropriateness of a given form is not likely to help the student understand where the processing has gone wrong. The potential of the computer for addressing the problems of individual language learners is only just beginning to be realized (see Garrett, 1987).

[9] Another important variable in the learning situation, the different challenge posed to native speakers of English by the learning of highly inflected vs. less highly inflected languages, of typologically similar vs. dissimilar languages, etc. has as yet received almost no explicit research attention in foreign language education. A very high proportion of serious work in second language acquisition theory and empirical research in interlanguage has been done on English, and of the work on the acquisition of other languages a high proportion has been done on French and Spanish, languages which are both closely related and relatively less inflected. It should not be assumed that second language acquisition theory which has been developed on the basis of this limited research -- still less any general notion about how much explicit attention should be paid to grammar -- is generalizable to all classroom language learning.

References

Byrnes, H. (1987). Proficiency as a framework for research in second language acquisition. *Modern Language Journal, 71*, 44-49.

Corder, S. P. (1977). 'Simple codes' and the source of the second language learner's initial heuristic hypothesis. *Studies in Second Language Acquisition*, 1-10.

Garrett, C. [same author as Garrett, N.] (1982). *In search of interlanguage: A study of second-language acquisition of German syntax*. Unpublished doctoral dissertation. University of Illinois, Urbana-Champaign.

Garrett, N. (1986). The problem with grammar: What kind can the language learner use? *Modern Language Journal, 70*, 133-148 (esp. Note 23).

Garrett, N. (1987). Psycholinguistic grammar and CALL [Computer-assisted language learning]. In W. F. Smith (Ed.), *Modern media*. Lincolnwood, IL: National Textbook.

Garrett, N. (in preparation). *Processing as the central construct in language acquisition: Implications for the classroom.*

Givón, T. (1979). From discourse to syntax: Grammar as a processing strategy. In T. Givón (Ed.), *Syntax and semantics*, 12: Discourse and syntax (81-112). New York: Academic Press.

Karmiloff-Smith, A. (1981). Language as a formal problem space for children. In W. Deutsch (Ed.), *The child's construction of language*. London: Academic Press.

Kelly, L. (1969). *25 centuries of language teaching*. Rowley, MA: Newbury House.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Spolsky, B. (1985). Formulating a theory of second language learning. *Studies in Second Language Acquisition, 7*, 269-288.

Valdman, A. (1980). Communicative ability and syllabus design for global foreign language courses. *Studies in Second Language Acquisition, 3*, 81-95.

VanPatten, B. (1986). The ACTFL proficiency guidelines: Implications for grammatical accuracy in the classroom? *Studies in Second Language Acquisition, 8*, 56-67.

WHAT PRICE GRAMMAR?

Simon Belasco
University of South Carolina

The following is a commentary on Nina Garrett's "The Role of Grammar in the Development of Communicative Ability." Many of her observations on grammar in present day foreign language instruction reflect keen insight into the problems affecting acquisition of a foreign language in the unicultural, artificial situation represented by the average FL classroom. Although I agree with her assessment of the problems, I feel that foreign language instruction has yet to benefit from any solution that has been proposed thus far from any source. The emphasis on production in foreign language learning -- which has characterized virtually every approach to foreign language instruction since Classical times--has never been successful. This "quick fix" philosophy has contributed to the persistent re-invention of the "square wheel." As long as foreign language educators continue to pursue this approach, the great mass of foreign language learners will be unlikely to acquire a second language in the classroom.

According to Dr. Garrett, there are three related fields of research providing information on how grammar works in second language learning: (1) foreign language education, (2) natural second language acquisition, and (3) first language acquisition. She believes that FL education and natural L2 acquisition have been unable to supply generally acceptable answers because they have not been asking the right questions. The teaching of grammar has been overly concerned with the acquisition and use of rules of morphology and syntax, which are erroneously supposed to be the basis for the learner's performance. The crux of the problem lies then in the Chomskyan construct of linguistic competence which--because of its central position--has interfered with our attempts to develop a coherent theory of grammar in second language acquisition. I shall not dwell on the point that certain foreign language educators -- myself among them, and not Chomsky--are responsible for attempting to apply his theory to second language acquisition. Chomsky has stated on several occasions that generative grammar has nothing to contribute to that area (Chomsky, 1966). He does claim, however, that his theory of Universal Grammar applies directly to the field of first language acquisition. Interestingly enough, Professor Garrett does say that "questions asked by first language acquisition research suggest important hypotheses which should be explored in our domain" (Garrett, 1987), which is precisely what Chomsky advocates. Suppose we accept the tenet that the basis for performance is the desire to express some meaning through a complex set of processing rules, i.e. sociolinguistic, pragmatic, and discourse rules of a language which "are derived from native speakers' intuitions about what is appropriate." Furthermore, we may easily accept the tenet that grammatical competence can be understood as a template "against which any utterance can be mentally matched." There still remains the problem of how all this is to be implemented in the classroom--that is, how can this serve to get our students to learn a foreign language?

The lesson we can learn from first language acquisition is that, despite the poverty of the stimulus, i.e., the relative paucity of "linguistic exposure," the child develops "a rich and highly articulated system of grammatical competence in a specific way" (Chomsky, 1981). This implies that the utilization of any contrived inductive base is highly suspect. The entire framework underlying the teaching materials in high schools and colleges today is geared to "producing" language before students "acquire" language. We try to measure proficiency before students are proficient. We say that listening comprehension should precede production, but we continue to emphasize production. We still teach grammar as we teach math: feminine noun, feminine adjective; plural noun, plural adjective--A is to B as C is to D. Even if our students mastered every exercise of every lesson in an entire course, they still would not attain the level of comprehension, linguistic, or communicative competence achieved by a three year old native speaker. We still consider grammar in terms of the same traditional grammatical principles which have appeared in foreign language textbooks for more than fifty years: indicative and subjunctive moods, comparison of adjectives and adverbs, regular and irregular verbs, etc.-- some 50 principles all of which are supposed to be internalized by the student after one or two years of study. There is little experimental justification for the choice of these principles or any principles for that matter. What is particularly distressing is that many foreign language teachers have internalized thoroughly all 50 principles and still are unable to understand the language when overhearing a conversation engaged in by native speakers--let alone comprehending a film without subtitles in the foreign language they teach (Belasco, 1985).

As Professor Garrett points out, much of the problem stems from the formal teaching of grammar, which is supposedly designed to help the language learner acquire linguistic competence. Although no empirical evidence shows that the kind of grammar taught in schools is the grammar internalized by a native speaker, neither is there sufficient evidence to show that the sociolinguistic, pragmatic, and discourse rules we may have in mind are actually representative of "native speakers' intuitions about what is appropriate." Adding rules or replacing old rules by new ones will not solve the acquisition problem. Production may be the ultimate goal in second language acquisition, but comprehension is crucial to communication. No matter how sophisticated the quality of the language learner's production, communication will break down if he fails to comprehend the speech of the addressee. Most foreign language educators do not seem to realize the import of this statement.

It is without doubt the major reason why foreign language learning in the classroom has been unsuccessful. Comprehension should be the primary candidate for intensive research in second language acquisition.

A place to begin is where "real language" occurs. What is the nature of the "highly structured grammar" that a child infers on the basis of limited and defective data? What fundamental principles "sharply restrict the class of attainable grammars and narrowly constrain their form, but with parameters that have to be fixed by experience" (Chomsky 1981)? "A theory of acquisition is therefore not one of testing hypotheses, but rather one of fixing parameters" (Bouchard, 1984). Whatever characterizes real language, it is "complete and simple" enough to enable the child to effect internalization within the constraints on "competence" and "appropriateness" imposed by his inherited language faculty. No pedagogical approach to second language acquisition has been able to duplicate what is the most salient characteristic of first language acquisition: the ability of human beings without hearing impairment to acquire a first language in a normal social context. In fact, one cannot prevent himself from acquiring a first language under normal physical and social conditions. Yet in the formal context of the ordinary classroom, despite many years of study, few language learners--if any--acquire a second language (Belasco, 1981a, 1981b).

What is often considered today as measurement of foreign language proficiency is better interpreted as measurement of foreign language achievement. There is nothing disparaging in measuring achievement. The attainment of a high level of listening comprehension in second language learning is possible and favors the acquisition of the other skills. It is not difficult to outline a simple--yet effective--way of attaining a good level of listening and reading comprehension, designed to reduce the formal presentation of grammar. We may start with relatively simple materials but asymptotically eliminate them in favor of "live," genuinely native materials.

In this approach, students--prior to coming to class--are required to listen to French-English materials presented in dovetailed fashion as in the excerpt below. They also have a printed text.

> A la charcuterie / At the pork-butcher shop. // Madame Neveu / Mrs. Neveu / doit préparer un dîner / has to prepare a dinner / pour quatre personnes / for four people. // Elle entre dans une charcuterie / She goes into a pork-butcher shop / pour acheter le hors d'oeuvre / to buy the hors d'oeuvre / et le plat principal / and the maindish //, etc. (French text based on Walz and Piriou, 1985).

Students read and listen to the passage until they can understand any excerpt extracted from the original text without further consulting the source language. When they come to class, students are asked not only to provide the meaning of selected French excerpts but are also required to respond in French to questions asked in French based on the content of the text. The grammar and the cultural material derive strictly from textual content. Grammatical discussions arise from student questions, student errors, or relevant points brought up by the teacher. Suppose in answering a question, a student has processed the material in such a way that he says: *Le charcutier donne Madame Neveu 2 tranches de jambon instead of Le charcutier donne 2 tranches de jambon à Madame Neveu. It is at this point that grammatical constraints are brought up for discussion. In a sentence such as "The pork-butcher gives 2 slices of ham to Mrs. Neveu," students learn that "2 slices of ham" bears the semantic role of theme and "to Mrs. Neveu" bears the semantic role of goal (for semantic roles of pedicate-argument structures, see Marantz, 1984). The student is informed that the parameters of English are fixed so that "in simplex sentences" the juxtaposition of verb plus [+animate] noun with the role of "goal" is permitted for certain marked verbs such as "give," "send," "tell," whereas the parameters of French do not permit such juxtaposition.[1] Thus English has "John says / speaks / tells / the words to Mary" and "John tells Mary the words"--but not "*John says / speaks / Mary the words." French, of course, permits only the analogue Jean dit les mots à Marie.

Other errors resulting from mis-processing are handled in the same way. For example, in treating certain hypotactic constructions in French, no pedagogical textbook differentiates between prepositions and complementizers introducing embedded clauses, thereby neglecting the constraints involving à, de, and "zero" determining the meaning of utterances containing embedded constructions. Thus, movement and replacement operations may require different pronominal forms according to whether a preposition or complementizer is used. For example, de is a preposition in Jean a rêvé de le faire "John dreamed of doing it," as pronominalization of the embedded clause reveals: Jean en a rêvé "John dreamed of it"). On the other hand, de is a complementizer in Jean a oublié de le faire "John forgot to do it" since pronominalization results in: Jean l'a oublié "John forgot it." The use of an incorrect complementizer form may result in faulty processing, obscuring "who does what to whom." Thus, using à in the sentence Ils lui ont demandé à partir "They asked (permission of) him to leave" signals matrix subject control, i.e., Ils is interpreted as the "leaver," whereas the use of de signals matrix object control in the sentence Ils lui ont demandé de partir "They asked him to leave," where lui is interpreted as the "leaver." Similarly, presence or absence of the complementizer de signals matrix object control in the sentence Elle lui a dit de le boire "She told him to drink it," but signals matrix subject control in Elle lui a dit l'avoir bu "She told him that she had drunk it."

Chomsky's recent version of the Extended Standard Theory, government-binding (GB) theory, has shifted the focus from the study of rule systems to the study of the system of principles. The concept of "base generation" has no clear meaning. Representations at different levels of structure are determined by fixing parameters of Universal Grammar, which constrain movement and coreference of arguments involving names, pronouns, anaphors, and other referential expressions such as quantifiers and variables. The conditions on anaphors and pronouns, for example, account for differences in meaning between reflexives and pronouns in utterances such as Nous nous plaignons, which is ambiguous on the readings "We are complaining" and "We pity each other." Under this analysis, forms that are reflexives or potentially reflexive such as me, te, se, nous, vous may be either reflexive or pronominal, i.e., they may participate in either meaning. On the other hand, se has only the reflexive meaning of "complain" as in Elle se plaint "She is complaining," and le has only the pronominal meaning of "pity" as in Elle le plaint "She pities him." All types of printed matter, radio and television broadcasts, etc. abound in such constructions. Traditional grammars make little or no mention of grammatical processes such as establishing coreference with anaphors and pronouns, marking for selectional features, i.e., [±animate],or the relevance of semantic roles. At best, pedagogical grammars treat surface structure and are more useful as reference material rather than core material for the foreign language curriculum. The factors of processing and interlanguage can play an important role in the FL curriculum, but the "results" that may be obtained from ongoing research in GB theory have more to contribute to a "comprehension oriented" strategy in L2 acquisition. The formal teaching of principles and parameters-- let alone the linguistic jargon--need not be incorporated into such an approach. In applying the principles of grammatical processing, however, it is of prime importance for the student who has committed an error due to faulty processing or induction to have access to the constraints that source and target language native speakers observe in processing their respective languages.

Note

[1] Garrett, 1986 appears to minimize the role of selectional features such as [±animate], [±human], etc. and semantic roles generally in such sentences as 1."I gave my brother the money," 2. "I gave the money to my brother," 3, "he was given the money," 4. "I gave the the chair a coat of paint," 5. "he told his mother what had happened," etc. She gives the impression that what is at issue is the inability of the native speaker to recognize marking of the indirect object in the first sentence (see above), that both direct and indirect objects are permitted to act as subject of a passive sentence in some languages (sentences 1, 2, 3) but not in others, that it is possible to use an inanimate indirect object immediately following a verb like "give," (sentence 4) "so animacy is not the concept which is actually being encoded by the form," and that in sentence 5 "mother" is not recognized as the indirect object.

Garrett's arguments anticipate faulty processing based on traditional grammatical principles. It is not clear that "direct" or "indirect" objects are what is at issue in sentences 1 through 5 above. Thus, in French, it is not true that indirect objects do not act as subject of a passive sentence--all claims to the contrary. For example, the indirect object in such structures as Le roi a pardonné à / a obéi à / a désobéi à / la reine may be passivized as La reine a été pardonnée / obéie / désobéie. It would appear then that an "object" to be passivized in French must bear the semantic role of "theme"--not bear the grammatical feature "direct." This is brought out more clearly in such sentences as L'avocat a mal conseillé à l'inculpé which may be passivized as L'inculpé a été mal conseillé because l'inculpé bears the semantic role of "theme," despite the fact that it functions as an "in-direct" object. Note that if an infinitive is embedded in the verb conseillé, the embedded infinitive assumes the semantic role of "theme" and l'inculpé that of "goal," thereby barring the latter from passivization: L'avocat a conseillé à l'inculpé de plaider non coupable, but not *L'inculpé a été conseillé de plaider non coupable. Moreover, in sentence 4 above, the issue is not that a verb such as "gave" may be followed immediately by an inanimate noun or not, but that a verb such as "told" but not "said" in sentence 5 may be followed immediately by an animate noun with the semantic role of "goal" in the same simplex sentence. Thus, "told" but not "said" qualifies for marking in the frame "he his mother what had happened." On the other hand, "said" may be followed immediately by an animate noun if it is in a complex sentence, i.e., a following embedded sentence: "he said (that) his mother knows what happened." In any event, clauses are neutral as to the features [±animate] or [±human], the latter being ascribed to lexical items that bear a referential index. The problem with processing in second language learning is not always visible in the learner's interlanguage but may be due to misinterpretation by the language analyst.

References

Belasco, S. (1981a). Aital cal aprene las lengas estrangièras, comprehension: The key to second language acquisition. In H. Winitz (Ed.), *The comprehension approach to foreign language instruction* (pp. 14-33). Rowley, MA: Newbury House.

Belasco, S. (1981b). Toward the acquisition of a second language nucleation revisited. In H. Winitz (Ed.), *Native language and foreign language acquisition* (pp. 309-323). New York Annals of the New York Academy of Sciences, Vol. 379.

Belasco, S. (1985). Toward the identification of a core grammar in L2 acquisition. [Review article of lectures on government and binding (The Pisa lectures) (1981b)]. *Studies in Second Language Acquisition, 7,* 91-98.

Bouchard, D. (1984). *On the content of empty categories.* Dordrecht: Foris.

Chomsky, N. (1966). Linguistic theory. In *Language teaching: Broader concepts. Reports of the working committee 1966 Northeast conference on the teaching of foreign languages* (pp. 43-44).

Chomsky, N. (1981a). Principles and parameters in syntactic theory. In N. Hornstein & D. Lightfoot (Eds.), *Explanation in linguistics, the logical problem of language acquisition* (pp. 32-75). London and New York: Longman.

Chomsky, N. (1981b). *Lectures on government and binding, the Pisa lectures.* Dordrecht: Foris.

Garrett, N. (1986). The problem with grammar: What kind can the language learner use? *Modern Language Journal, 70,* 133-148.

Garrett, N. (1987). The role of grammar in the development of communicative ability. In *Symposium on the Evaluation of Foreign Language Proficiency.* This volume.

Walz, J., & Piriou, J. P. (1985). *Rapports, language, culture, communication.* Lexington, MA: D.C. Heath.

A RESPONSE TO THE ROLE OF GRAMMAR IN THE DEVELOPMENT OF COMMUNICATIVE ABILITY

Theodore V. Higgs
San Diego State University

If this paper had its own title, it might be called "An Agenda for a National Debate on the Role of Grammar in Adult Foreign-Language Instructional Programs." Fortunately, the record will read simply that it is a response to Dr. Garrett's paper, "The Role of Grammar in the Development of Communicative Ability."

Garrett has said that, "Grammar may fairly lay claim to having generated the most controversy between proponents and opponents of oral proficiency testing and the use of the proficiency guidelines as a framework either for testing or for teaching and syllabus design." I find it to the point (which is also undefined) or to another's position. As Garrett points out, "For grammar, the debate is whether we should be teaching it at all."

Debating whether grammar should or should not be taught is possible only when the putative debaters are talking through each other, with neither side understanding nor perhaps even listening to the other. In the context of this symposium, it seems appropriate that we take more than a casual look at what might be understood by the term "grammar," in order to diminish the heat and increase the light. If there is to be a debate about the role of grammar in adult foreign language acquisition, it can be undertaken seriously only if the participants at least agree in advance on what is being debated. In short, we need to agree on the agenda.

There are many senses in which it is legitimate to invoke the term "grammar." As I was preparing this presentation, I asked Mike, the technical supervisor of our language laboratory (and a certifiable computer expert), what he understood by the word "grammar," and what kinds of images the word evoked in his mind. He replied that for him "grammar" meant the rules and formulae that account for the way sentences are supposed to be constructed. The mental images that the term conjured up for him included long, convoluted rule statements, charts, and hopelessly opaque terminology that had no meaning for him at all. When I asked him if he had studied foreign language in college, he went into an extended monologue about flunking first semester German three or four times, then switching to French with the same result, and finally changing universities, in favor of one that had no foreign-language requirement for graduation.

Mike's response to my question about the meaning of the word "grammar" was remarkably close to the first sense given in *The Random House College Dictionary* under the entry "grammar." That is, "the study of the formal features of a language, as the sounds, morphemes, words, or sentences." This perception of grammar--we might call it the cookbook perspective--is the source of the arcane metalanguage that most linguists and many foreign-language teachers hold sacred. It is from this source, that the rule to account for English sentences such as "I want him to go" comes from, i.e., "the subject of an infinitive clause is always in the accusative case." What matter that English most likely has no case system? It's a wonderful sounding rule, it's something to "teach" your English students, and for some native speakers of English, at least, the ability to recite such a "rule" established solidarity with other like minded individuals.

Equally wonderful sounding are some of the "rules" found in foreign-language textbooks. Try, for example, grammar point #52, entitled "Subjunctive After Nonexistent and Indefinite Antecedents:" "In Spanish, when the antecedent of an adjective clause refers to someone (something, someplace, and so on) that does not exist, the subjunctive must be used in the adjective clause."[1] How would you motivate your students to master such a rule for productive communication?

My point is not that such grammatical statements are valid or invalid in any absolute sense. I trust, however, that no one could seriously argue that they facilitate either the learning or the acquisition of a foreign language. In fact, anyone who would advocate a foreign-language pedagogy in which the use of such a perspective on grammar is perceived as crucial or even significantly facilitative, must accept the responsibility of defeating in open debate the proposition that mastering a formal grammar couched in elegant metalanguage requires that the foreign-language student learn two languages, not just one, with little or nothing to suggest that the two learning tasks are productively related.

Requiring that a student first learn such facts about the language as a prerequisite to learning the language is rather like requiring that a potential vacationer first memorize the road map he may one day have to follow, should he actually decide to take a vacation. In short, why double the learners' task? Any debate about the role of grammar in adult foreign-language acquisition that proceeds from this perspective on "grammar," if it gets started at all it will be very short-lived.

The same article on the word "grammar" in *The Random House College Dictionary* notes farther on in sense five, that grammar is "knowledge or usage of the preferred or prescribed forms in speaking or writing." This perspective is one that deserves a longer look from members of our profession, for irrespective of our pedagogical proclivities, all of us aspire to helping our students master the usage of the preferred or prescribed forms of the languages we teach. Such an aspiration is not born out of straight-jacket prescriptivism. Rather, it implicitly recognizes that the purpose to which language use is best put is to share and exchange information confidently. As I have suggested elsewhere (Higgs, 1985a), a grammar is best perceived as a system by means of which messages are converted into language.

The messages we choose to transmit through language can serve virtually an infinity of functions: to inform or misinform; to protect ourselves; to establish and preserve in-group relationships or to exclude unwanted out-group members; to speculate, theorize, or hypothesize; to negotiate, persuade, counsel or advise; to offend or to sooth an offended party. Such message types or "communicative functions" are ultimately enabled and supported by successfully combining appropriate lexis with appropriate grammatical structures in a given language.

The business of foreign-language teachers is that of enabling their students to receive and produce meaningful messages. The purpose of the foreign-language classroom is to provide a sheltered environment in which students will be motivated to begin receiving and producing meaningful messages. A major responsibility of the foreign-language instructor is to help the students correctly identify for themselves the object of study and a productive learning strategy for mastering it. In order to meet this responsibility, the instructor must first identify and then deal with--either by encouraging or defeating--the perceptions that the students bring to their task.

Over the last several years, early on in my elementary Spanish classes I have elicited self reports from my students regarding their own perceptions of foreign-language learning. That is, I have asked them to tell me in a paragraph or two what it is they think they are being asked to learn, and how they think they should go about learning it.

The overwhelming perception that has emerged from this informal research project is that students think that Spanish is the same as English, except that it involves a different vocabulary (Higgs, 1979). Therefore, from the students' perspective, learning Spanish consists mostly in learning Spanish vocabulary, and stringing it together by implicitly following the norms of English syntax. This is a tempting hypothesis, and one which some pedagogies implicitly espouse, either by distorting Spanish structures or by suppressing other structures that differ markedly when compared to the native-language structure. Unfortunately, a minimum of exposure to Spanish shows students almost immediately that a simple process of algorithmic relexification fails to produce the desired result, which of course is a successful transmission or reception of a message. Literally on the first day of instruction, for example, students learn that the Spanish utterance Me llamo Ted 'My name is Ted' contains no word that means "my," none that means "name" and none that means "is."

Even more important than the failure of the "lexical analog hypothesis," however, is the fact that in its surface structures Spanish is not an SVO language. Thus, a lexical strategy for understanding spoken or written Spanish will fail to identify the cognitive synonymy of Carlos le sigue a María and A María le sigue Carlos 'Carlos follows María.' It will at the same time fail to distinguish between A María le sigue Carlos 'Carlos follows María' and María le sigue a Carlos 'María follows Carlos.'

The fact of the matter is that because in Spanish surface-structure morphology and syntax carry a heavier functional load, and hence a greater communicative burden, than is the case for English, its grammar allows a more flexible constituent ordering to carry certain messages. That is, in English, flexible constituent ordering such as is found in pairs like "I gave Mary the book" and "I gave the book to Mary" is more the exception than the rule.

A grammar of Spanish that merely points out that word or constituent ordering is flexible, however, loses the opportunity to show how the grammar of Spanish transforms meaning into language. Students typically understand "flexible" or "optional" to mean that it doesn't make any difference how you do something; that one "option" is equivalent to another. Consider, for example, the "subject-of" constituent in Spanish. Students' communicative needs, whether for reception or production, are not well served by a "rule" that says that the overt expression of the subject of a sentence is "optional." Such a statement entirely ignores both the role of context interpretation in communication, and implicitly defines "grammar" as a sentence-level phenomenon in which only "conditions on wellformedness" are important.

Pedagogically there are many interesting and helpful observations to make with respect to the subject constituent, not only as they impact on its overt presence or absence in a given communicative environment, but also in terms of its placement relative to the verb. In no interesting sense is either the overt mention or omission of a subject or its placement in a sentence "optional." For any given communicative context, it is either obligatorily mentioned or obligatorily omitted, and if mentioned is either placed obligatorily before or obligatorily after the verb. Most importantly, the reasons for mentioning the subject constituent or not, and for placing it wherever one places it, respond to clear semantic differences that even beginning learners can understand. When the communicative impact of a structure is clear to the students they can indeed "attend to meaning first" and the overall comprehensibility of a speech sample is improved (Higgs, 1985b; Krashen, 1982).

Pointing out the communicative impact of subject use and placement versus subject omission can be done without resorting to even minimal use of metalanguage. It consists merely of making overt something that may be being ignored or misinterpreted by the recipient of the input. A pedagogy that confuses teaching formal grammar with teaching overt grammar misses the opportunity to contribute both to learning and to acquisition, even assuming that the difference between such terms is real and important. The student who does not know that there is a difference in meaning between Yo hablo español and Hablo español will neither learn nor acquire the competence to express or perceive such a difference, even though he might produce well-formed sentences. Even more important, no amount of comprehensible input will inform this lack of awareness, simply because the salient portion of the message is not comprehended.

It is entirely unrealistic to debate whether or not "grammar" should be taught in adult foreign-language classes, as long as we understand that the grammar is the device or the system though which we convert meaning into language. Under this perception of "grammar" it is impossible not to teach it. The only worthwhile debate is how best to teach grammar. The profession's needs are not served by hyperbole or gratuitous misinterpretations of possible positions with respect to the teaching of grammar. It is folly, for example, to attack someone's recognition of the crucial role that accuracy plays in successful, that is correct, appropriate and confident communication, and their advocacy of teaching for accuracy as "schizoid" (i.e., presumably claiming to teach for "communication" while actually teaching "grammar") or as implicitly entailing a fixation on discrete point teaching or testing.

Further, there is no basis to conclude a priori that someone who is teaching grammar is doing so to the exclusion of any other type of activity, and even less basis to assume that he or she is teaching traditional, formal grammar.

The purpose of instruction is to inform and enlighten, not to obfuscate and confuse. A fundamental premise of any instructional endeavor is that the instructor starts with something that his students already know, in order to move them towards the current object of study. No one would argue that the typical American adolescent approaches foreign-language study with ready-made metalanguage to account for the functioning of his own native language. On the contrary, native speakers of any language are notoriously unaware of how their language works, to say nothing of their possessing an elegant elaboration and explanation of its elements and processes.

I have suggested elsewhere (Higgs, 1985a) that all classroom behavior counts as "teaching," and that "teaching grammar" means making as clear as possible the relationships that link meaning and form by tying target language forms to meanings or other cognitive constructs the students already possess. In the same article I urge, and I am quoting here, that teaching grammar "explicitly does not mean teaching formal rules, either of the sort associated with theoretical linguistics or with traditional grammar." The essential message of teaching grammar is a blend of "If you choose to mean this, say that;" with "if you say this, you mean that, whether you wanted to or not." Such a position with respect to the teaching of grammar is entirely consonant with Krashen's assertion (1982) that "we acquire by 'going for meaning' first, and as a result, we acquire structure."

If students are indeed to go for the meaning first, a certain amount of "interventionist pedagogy" is a necessary part of foreign-language teaching. Languages are, as structuralist linguists delighted in pointing out, fundamentally arbitrary and conventional. They are "arbitrary" in that there is no essential connection between a given form and its meaning. They are "conventional" in that they work the way they work only because an interesting majority of their speakers agrees that that's the way they work.

Students are accustomed to treating college-level courses as cognitive or intellectual problem-solving exercises. They expect to be given the "formula" and then to apply it to new and previously unexperienced situations. But given the arbitrary and conventional nature of natural language, no amount of intelligence or application will make it possible to "intuit" a lexical item or grammatical form in a foreign language. For example, one cannot tell "by looking" at a linguistic form or by hearing it whether it will turn out to be regular or irregular. Even more important, not even the meanings that the target language encodes are necessarily part of the students' consciousness. That is, certain meanings or relationships will be "salient" and hence overtly marked in the target language and not in the native language. Thus, often times in the foreign-language classroom it is often crucial to focus first on the semantic distinctions that are being signalled in the target language before any attention at all is given over to the forms that signal them.

Continuing in the vein of interventionist pedagogy, it seems quite plain to me, counterclaims notwithstanding, that the provision of feedback and correction must be an integral part of foreign-language instruction. What is important is to recognize that the purpose of this type of formative feedback is not to ensure somehow that the corrected error will never recur. Rather it is to let the student know what the impact of his communicative effort is likely to have been. The purpose of correction is to provide feedback in two forms: whether the message can be successfully received, i.e., whether it has been successfully transmitted, and what additional information has been transmitted along with the intended message. Often an unintended adjunct to the message at the reception end of the line is something like, "I hope I never have to depend on this person to communicate anything important." Stated simply, one of the goals of any communicative effort is not only to make oneself understood, but to ensure that one will not be misunderstood.

In addition to her comments on grammar and its role in the adult foreign-language program, another matter addressed by Garrett is whether the Proficiency Guidelines can or should serve as a framework for teaching and syllabus design. I should first make clear that it is not my perception that the Guidelines claim to represent a theory of first or second language acquisition, nor do they claim that other things being equal, individuals whether in a formal or an informal environment progress systematically in a linear fashion through the levels as characterized by the level descriptions.

The levels ultimately derive from a needs assessment carried out some thirty years ago that attempted to identify the types of communication and assumed foreign-language proficiency required to perform certain types of work in US government installations abroad. The fact that some people perceive a kind of developmental sequence in the Guidelines is interesting and is worthy of field research, but is not to be taken as a given. The Guidelines themselves, however, are explicitly pedagogy independent and in my opinion inform neither methodology nor syllabus decisions.

In terms of the pedagogical inferences that one might draw from the Guidelines, it is important to note that the techniques of evaluating and rating a speech sample recognize "concept control, partial control and full control" over the communicative functions and content areas that the OPI purports to sample. This implies that to the extent that the Guidelines may inform decisions regarding syllabus design, one must not assume that functions and content areas associable with a given proficiency level should be suppressed until all the functions and content areas of a presumably lower level have been mastered. Nor does it assume that all linguistic features associable with a given proficiency level must or even should be explicitly taught.

Depending on the target language, certain elements of the expression system are relatively more difficult to master than are others. From one perspective, it seems reasonable to me to argue that of those elements that are deemed essential for predictable communicative needs, the more difficult they are to master, the earlier they should be introduced into the syllabus, on the argument that more time and exposure is necessary to facilitate their incorporation into the students' nascent grammar. From another viewpoint, however, it seems just as reasonable, to argue that in an elementary foreign-language program one should present material that is easiest to manage.

Deciding between these two possibilities is a matter of policy, which can be made only in the context of a real program involving real students. Among the variables that would impact on such a decision would be the age of the client population, the perception of the program's intent (e.g., to provide in three or four semesters the presumed entire exposure that the client population will have to the target language, or to start them off, on the assumption that they will continue their study and exposure for a much longer period of time.)

In this context, even the expression "elementary courses" is ambiguous. If elementary Spanish, for example, means the ability to handle concrete, predictable communication involving the normal intercourse among peers, then certainly the ability to narrate and describe in past, present and future time, and to commiserate over a friend's misfortunes are elementary communicative needs. Yet structurally, they entail distinguishing between perfective and imperfective past and using the subjunctive mood, both famous learning/acquisition problems for anglophones. If an "elementary" program is designed to provide for these communicative needs, relatively "difficult" material must be introduced early on in the syllabus in order for there to be time for its assimilation.

On the other hand, if elementary Spanish means a level of Spanish that promises not to be too difficult, and in which the students' natural attraction to the lexical analog hypothesis is not to be seriously threatened nor even challenged, then the conceptually difficult material might reasonably be deferred until more advanced study. It seems fairly clear to me that the existence of the Guidelines will not inform such a decision. In short, it seems that both matters of syllabus design and pedagogical orientation should be taken up independently of the Guidelines.

In his 1980 paper on "Communicative Ability and Syllabus Design for Global Foreign Language Courses" Valdman (1980) suggests reducing the number of structures to be learned in beginning courses, on the assumption that the goal of such courses is to "impart minimal communicative ability at the elementary/intermediate levels." Provided that the structures to be excluded are excluded for the right kinds of reasons, I agree with this recommendation. The problem lies in identifying just what structures are necessary to achieve the stated goal. I suggest that one criterion that can be applied in deciding which structures to prune from a beginning course is that of the "overtness" or "transparency" of the relationship between the forms to be taught and the possible meanings such forms convey. On the one hand, we must be careful not to prune away too much, arguing that only what can be completely acquired during the program of study should be presented. On the other hand, however, we must not assume that an elementary or introductory foreign-language program should exhaustively sample the full arsenal of structural weapons that the target language might provide.

Note

[1] From a current college-level Spanish textbook that shall remain nameless.

References

Higgs, T. V. (1981). Some pre-methodological consideration in foreign language teaching. *Canadian Modern Language Review, 37*, 309-319. (Reprinted from *Modern Language Journal*, 1979, 63, 335-341)

Higgs, T. V. (1985a). Language acquisition and language learning: A plea for syncretism. *Modern Languge Journal, 69*, 8-14.

Higgs, T. V. (1985b). The input hypothesis: An inside look. *Foreign Language Annals, 10*, 197-203.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. New York: Pergamon.

Stein, J. (Ed.). (1982). *The Random House college dictionary* (revised ed.). New York: Random House.

Valdman, A. (1980). Communicative ability and syllabus design for global foreign language courses. *Studies in Second Language Acquisition, 3*, 81-96.

Summary of the discussion session on THE ROLE OF GRAMMAR IN THE DEVELOPMENT OF COMMUNICATIVE ABILITY

Laurie Knox
Indiana University

The discussion opened with a question addressed to Nina Garrett about the extent to which grammatical accuracy should be required of language users beyond productive and receptive comprehensibility. Reference was made to two opposing views on this question: Vicky Galloway's view of grammar as a meaning access system, and Alice Omaggio's view, which seemed to the questioner to treat grammatical accuracy as an intrinsic value. The initial reply was that because of the many variables involved in comprehensibility, the question was too general: one cannot specify globally an acceptable level of accuracy for every grammatical form at every level of learning. Participants discussed some of the variables involved in the relationship between comprehensibility and accuracy. It was noted that languages vary in the degree of accuracy which is required to achieve comprehensibility: highly inflected languages require a greater degree of comprehensibility than less highly inflected languages. Also, a distinction was made between grammatical features which are essential to conveying meaning, such as case-marking particles in Tagalog, and grammatical features which are not essential, such as gender markers in German. Reference was made here to Albert Valdman's notion of "grammatical etiquette."

Pragmatic and sociolinguistic variables in the relationship between accuracy and comprehensibility were also discussed. One participant commented on research showing that a hearer's familiarity with non-native speakers is an important variable influencing comprehensibility. Another participant asserted that while comprehensibility is easy, and may not require sophisticated use of language, effective communication, when defined as the ability to bring about action on the part of hearers, requires skills beyond mere comprehensibility. It was argued that the concept of comprehension as a goal of accuracy is not adequate to account for the many sociolinguistic variables involved in the relationship between accuracy and communication. The French waiter who refuses to serve a customer who misuses his language, even though he understands him, was given as an example. Other participants agreed that while comprehensibility, defined as intelligibility, is a minimal acceptable level of accuracy, it should not be the target for proficiency evaluations or the goal of language programs.

From this point, the discussion moved to the question of what kind of accuracy language programs should choose as a goal for their students, and how goals of accuracy should be achieved. Some participants viewed this question as an issue of pedagogical policy which individual teachers and programs should decide on the basis of their specific needs, while others addressed the issue from a theoretical point of view. One participant expressed concern that the term "successful language learning," which seemed to him to be a central issue in debates about accuracy, is normally defined externally to the learner. A suggestion was made that learners be allowed to define success in language learning for themselves.

A participant argued that the question which most people ask about a language program is "When are they going to learn to talk?" and that, given this goal, accuracy must be secondary to the primary goal of communication in language classrooms. In support of this point of view, a participant made a proposal for research about the effectiveness of a course in German in which accurate use of the forms of articles, which, it was argued, are not essential to meaning, would not be required for the first year of study, and instead, communication skills would be strengthened. Another participant criticised this proposal with the observation that structuring a syllabus based on the distinction between forms which are essential to meaning and those which are not would be impossible because of the redundancy of forms: in German, the articles which mark gender, a feature which is not essential to meaning, also mark case, which is essential to meaning.

Another participant argued that, rather than choosing the grammatical forms to be taught on the basis of how essential they are to production, teachers should choose to teach a receptive grammar first. Commenting on Simon Belasco's presentation, this participant argued that, given the evidence from second language acquisiton research that production of grammatical forms comes much later than reception, students' comprehension of input, rather than their production of errors, should be the source of knowledge about grammatical processing rules and which grammatical features should be taught.

Practical and theoretical concerns were related to each other in a participant's suggestion that the immediate goals of particular programs should be devised with consideration for learners' possible long term goals of achieving higher levels of proficiency. An argument was made that immediately attractive goals of achieving particular levels of communicative ability in particular settings may defeat long term goals by reinforcing incorrect processing rules. An example was given of students of German who had learned the language informally, and who, later, were unable to go beyond a 2+/3 proficiency level, despite rigorous study. Policy makers, it was asserted, should consider the theoretical question raised by Garrett's paper of how to guide processing to prevent fossilization.

Another issue raised in the session concerned the role of what Theodore Higgs, in his presentation, called "the metalanguage" of grammar: the complicated rules and traditional grammatical terms taught in many foreign language classes. One participant observed that policy-makers and politicians, influenced by pop-psycholinguists, see the learning of this metalanguage as a principle reason for studying a foreign language. Other participants agreed that the question of metalanguage must be considered in light of a broader issue about the goals of foreign language education, goals which include, beyound second language learning, improving students' understanding of their own language and enhancing their acceptance of cultural differences. One participant remarked that if policy makers want metalanguage, professionals and academics concerned with language acquisition have a responsibility to provide input at the policy level to make sure that the metalanguage being taught is the best possible metalanguage, based on their professional knowledge. Another participant remarked that the enemy in making metalanguage part of an educational policy is not the public, but instead the very group of professionals represented at the symposium: teachers and administrators who want concentrated grammar lessons in textbooks, and textbook writers and publishers who respond to their demands.

A third issue touched on in the course of the discussion focussed on the theoretical foundations of the proficiency movement. This discussion began with a question addressed to Nina Garrett concerning the theoretical basis of her own model. Garrett replied that, while the Chomskian concept of grammatical competence plays a role in her model, a competency theory is not sufficient for describing acquisition, since competency is what a learner has after acquiring a language, and not what he uses to acquire it. Garrett stated that her work was an effort to develop a model of performance which might be useful for describing language learning processes. At this point, one participant noted the diversity of theoretical perspectives in the proficiency movement and voiced concern that without theoretical unity the movement might become a short-lived bandwagon, similar to other movements in the field which have generated great excitement initially, but have not been sustained. Another participant suggested that the tradition of study represented by Dell Hymes and some British linguists might be a relevant theoretical foundation for the movement, because this tradition shares with the proficiency a primary interest in the nature of effective communication.

The session closed with a short discussion of how Nina Garrett's processing model might be implemented by the classroom teacher. The problem of teaching the relationship between the notion of the beneficiary and the grammatical form of the dative case was used as an example. It was suggested that the processing rules by which students link the notion of the beneficiary and the dative form are often misguided and lead to errors. Highlighting either the notion or the form in structured input is insufficient to guide students to accurate processing. Teachers need to become aware of the processing rules which students are using and to guide them systematically toward the development of more accurate processing rules.

In conclusion, the discussion in this session dealt with a number of topics concerning the role of grammar in the development of language proficiency, including the relationship between accuracy and comprehensibility, the appropriate goals of grammatical accuracy for language programs, the role of the metalanguage of grammar in the classroom, and the theoretical foundations of the proficiency movement.

L2 VOCABULARY ACQUISITION

Susan M. Gass
Michigan State University

This paper will consist primarily of a review of recent literature on vocabulary acquisition. In so doing, I will approach the issue of the lexicon from an acquisition perspective as opposed to a pedagogical one.

1. The centrality of the lexicon

The centrality of the lexicon in acquisition or in any form of language use is not in dispute. In terms of recoverability of information, inappropriate lexical use often results in an inability by an interlocutor to recover a speaker's intended message. When a non-native speaker makes grammatical errors, the information regarding the speaker's probable intent is recoverable. For example, if a learner produces the utterance in 1,

1. Can you tell me where is the train station?

the question will undoubtedly receive an appropriate response. If one makes a phonological error, the information is possibly recoverable from other information in the sentence or possibly from information beyond the sentence. For example, in the following exchange:

2. José: I have a problem.
Teacher: What is it?
José: It's about the urinals
Teacher: What about them?
José: I don't know what to do with them.

Despite the fact that there was a definite phonological problem, the teacher (a woman) was finally able to figure out that what José was talking about was not urinals, but journals and that he had not understood the journal assignment she had given. Or, similarly when a native speaker of Arabic said 'a virgin of English', listeners, because of the context, were able to figure out that he meant 'version of English'.

However, the lexicon is different, since often a lexical item may relay information which is directly counter to one's intentions. This may happen most frequently when the source of the lexical error is the native language, as may be the case with false cognates. Gairns and Redman (1986) give the following example uttered by a Spanish speaker:

3. I feel sorry for people who live in the suburbs.

Does the speaker think the suburbs are boring and middle-class, or ugly and poverty-stricken? This latter possibility is the most likely given a meaning of the translation equivalent suburbio 'slum quarter, shantytown'. A similar example comes from a native speaker of English who, in a faculty meeting of a Spanish department said:

> 4. No lo hemos hecho antes porque no hemos tenido el patrón para ese tipo de actividad [we didn't do it before because we didn't have the pattern for that type of activity].

However, patrón means not only pattern, but 'big boss, godfather'. She was immediately attacked for her "classist" remark.

The first two types are more likely to result in nonunderstandings which Varonis and Gass (1985) have defined as "those exchanges in which there is some overt indication that understanding between participants has not been complete," whereas the latter (lexical) may result in misunderstandings which may never be reconciled. The interlocutor in the lexical case (example 3) may only be left with a rather odd impression of this Spanish speaker.

Clearly, the examples are oversimplified since in reality there is a complex interaction of factors contributing to the issue of comprehensibility. Furthermore, it is often not possible to distinguish between a lexical and a phonological error. What may have a phonological source may not be interpreted in that way. For example, in 4, a student has gone to his professor to discuss admission to the Ph.D. program. He asked the following question:

> 5. What tragedies do I have to follow?

The professor lapsed into silence, apparently never realizing that the student had intended 'strategy' rather than 'tragedy'. As Hatch (1983, p.74) claims:

> When our first goal is communication, when we have little of the new language at our command, it is the lexicon that is crucial. ...The words...will make basic communication possible. [emphasis added]

Politzer (1978) (cited in Levenston, 1979) found that native speakers of German believed that vocabulary errors were the most serious of all error types. In that study native German speakers were asked to grade the seriousness of errors according to one of six categories (included were morphology, grammar, lexis and phonology). The lexical category resulted in the most serious error. Politzer concluded by saying that native speakers have intuitive knowledge of the importance of using the correct/appropriate word in communicating.

Despite the obvious role of the lexicon, it has had a unique place in second language research. Its uniqueness can best be characterized by lack of focus. It is not the case that vocabulary is not discussed, but it is more often discussed as a side issue rather than as the major topic of research. In fact, Sharwood Smith (1984) claims that it is misleading to think of the lexicon as the 'neglected' component since many areas of research impinge on questions of the lexicon. For example, in discussions of communication strategies, Tarone (1983) lists a number of strategies which are basically lexical in nature: 1) word coinage where the learner creates a new word to communicate a concept (airball for balloon), 2) certain types of borrowing, such as literal translation where the learner does a word for word translation from the native language ('He invites him to drink' for 'They toast one another'), 3) use of a native language term without translation (tirtil for 'caterpillar'). A second example comes from Kellerman (1979) whose main concern is with the use of native language information in a second language learning situation. I will discuss his work in greater detail below. However, what is important is that he uses the lexicon as a means of illustrating broader principles of second language acquisition. A third example of peripheral lexical concerns can be seen in work by Hakuta (1974, 1976) on prefabricated patterns. A final preliminary comment in this section on the importance and centrality of the lexicon has to do with issues of the separability of the lexicon from other components of the grammar. If we consider an alternation such as profound/profundity, it is not clear whether the alternation is best accounted for by means of phonology, morphology, word formation, lexis or a combination of these possibilities.

2. Second language acquisition and linguistic theory

From an acquisition perspective, or at least my perspective on acquisition, there are at least two reasons for studying the second language lexicon. First, and most obvious, is that knowledge about how a second language lexicon is learned informs us about the nature of second language learning. Second, how learners organize and use their lexicon informs us about the nature of language in general. With regard to child language acquisition and syntax, Beckwith, Rispoli and Bloom (1984) argue:

> descriptions of child language and syntactic theory are only adequate to the extent that they inform our knowledge of the mental representation of language. Just as the explanatory potential of linguistic theory is grounded in its psychology, so also must the explanatory potential of any child language theory be predicated by the grammar that the hypothetical learner can learn.

Despite the fact that this statement was addressed to deal with child language data, it seems consistent with claims of second language in that acquisition research (whether first or second) and a theory of language must be responsive to one another. A comprehensive theory of language cannot ignore the fact that humans learn language; conversely, theories of acquisition cannot ignore the fact that acquisition must fit into a more general theory of language.

3. Interconnections in the lexicon

I begin this discussion of the learning of the lexicon with a concept which has been around in a language learning context for a quarter of a century--nucleation. It is a term brought in to the language learning arena by Pike in 1960 and developed by Belasco in a 1965 article. To explain the concept I present a quote from Pike cited in Belasco.

> When a droplet is condensed out of a gas, or when a crystalline solid is precipitated out of a liquid, nucleation has occurred. Nucleation is involved in the first small clustering of atoms or molecules--say the

> first two or three dozen--into a structural pattern which will then be extensively duplicated in a repetitive pattern to form a crystal. It is difficult to get these first molecules to clump together...Yet once nucleation has begun, growth may proceed with great rapidity...Thus it appears that the energy barrier opposing nucleation is much greater than that opposing growth. The initial formation of starting nuclei is very difficult. The growth of these nuclei into larger units is very rapid and relatively easy and simple.

With regard to language learning, Pike claims:

> Some persons have memorized long lists of vocabulary items, and even extensive rules of grammar, without being able to speak the language. One might say that their learning is in a supersaturated condition, without nucleation. That is, though they may have many of the elements necessary for a conversation, they cannot in fact handle these. Specifically, they lack the structure, the "crystallization"--which gives a characteristic patterning to sentences and conversations.

The interesting aspect of this, from my perspective, is what he calls crystallization which I interpret to mean integration or interrelation of linguistic information within a learner's grammar. As Pike says, learners memorize long lists of vocabulary items, often with little interconnecting structure. While it is now commonplace to think of grammar (in the sense of syntax) as something which is interconnected and internalized, the lexicon seems to still be something isolated from other parts of our grammar. Even though a significant amount of instructional time may be devoted to presenting, explaining and defining terms, vocabulary instruction is often haphazard, arising without preplanning and regardless of the activity (Gairns & Redman, 1986). That is, it is rarely the case that lessons are organized to focus specifically on vocabulary. Furthermore, in instances of specific lexical instruction, it is often the case that we deal with isolated words or, at best, words embedded within a sentence.

A basic issue in research on the lexicon is what it means to know a word. Meara (1978) claims that it is "best described in behavioural terms as the ability to react to a word in ways which are considered appropriate by the speech community" [emphasis added]. By reaction he refers to such activities as understanding words against a background of noise or reading single words with minimum time exposure. However, even if we are to consider reactions, it is not necessarily the case that we, either as learners or native speakers, react to all words in the same way. I have argued elsewhere (Gass, in press) that language information can be conceptualized as a continuum ranging from "strongly" to "loosely" represented knowledge. This is clearly the case with L2 vocabulary. Some words are undoubtedly reacted to as a NS would, while others are not. Similarly, in retrieval there are L2 words which can be readily retrieved while others cannot. In both instances the former are "strongly" represented whereas the others are "loosely" represented. As Palmberg (1987) points out frequency and practice may play a large part in determining which items are most readily accessible.

While it may be the case that reactions, as Meara claims, are part of what is involved in "knowing" a word, I would argue that another important aspect is the establishment of appropriate lexical networks or features. That is, not only is language use, one aspect of which is reaction, relevant but language representation must also be considered a crucial part of what it means to know a word.

Telchrow (1982) discusses various ways of conceptualizing knowledge of the foreign/second language lexicon. A major concern in her article is with differences between productive and receptive knowledge and, importantly, which precedes which. The studies which she reviews reveal conflicting results. It is clear that learners do not have "two" distinct sets of vocabularies, which is to some extent what is implied in a dichotomy such as receptive versus productive. Rather, she proposes that vocabulary knowledge can be best represented as a continuum with the initial stage being recognition and the final stage being production. Production, in turn, is also not conceptualized in a monolithic fashion. Productive knowledge may involve producing a range of meanings, producing appropriate collocations, etc.

Yet another distinction has been noted by Bialystok and Sharwood Smith (1985). They refer to two aspects of language, "knowledge" and "control." Knowledge representation refers "to the way in which the language system is represented in the mind of the learner (the categories and relationships in long-term memory)" whereas control is "the processing system for controlling that knowledge during actual performance." The analogy which they make is to a library. The mental representation of language (knowledge) can be compared to the books of the library. Not only are there a number of volumes, but they are organized and interrelated in systematic ways. With regard to control, or access, the language user or library user needs to know how to get information which includes where to find the books (language information) and how to get at that information in the most efficient way possible. Their distinction between knowledge and control is useful with regard to vocabulary, particularly since it crosses the boundaries of the more traditional notions of productive and receptive knowledge. Both must include information regarding knowledge and control.

Another area of concern in considering interrelationships within the lexicon is that of word association. In 1978 Meara reported on a study of native English speakers learning French. In first considering data from native speakers, Meara claims that mental lexica are organized primarily along semantic lines. That is, in word association tasks, a native speaker is likely to come up with words which are similar in meaning to a stimulus word. In addition to similarity in meaning, adults are more likely to come up with paradigmatic associations as opposed to syntagmatic ones. Paradigmatic associations are those in which the response word is from the same part of speech; syntagmatic associations are those which complete a phrase. An example of the first type is given in 6 and an example of the second type is given in 7.

In other words, for adult native speakers of a language, words which have similar meanings are stored in such a way as to evoke one another. Thus, semantic networks are well-established in native speaker lexica.

For learners, the situation is different. The results from Meara's study suggest that semantic organization is less established in learners lexica. There appear to be conflicting principles of organization for learners, one being semantic and the other word form.

6. Paradigmatic associations (from Meara, 1978)

Stimulus	Response
man	woman dog boy child
hand	foot finger glove arm

7. Syntagmatic associations (from Meara 1978)

Stimulus	Response
brush	teeth
hold	hands
black	mark
bank	robber

Below is an example of native French and L2 French responses to the stimulus word profond 'deep'.

8. From Meara (1978)

Stimulus	Native	Learner responses
profond 'deep'	puits 'well'	plafond 'ceiling' (prendre 'to take') (professeur 'professor')

What is particularly interesting from the results of his findings is the implications he suggests for learner difficulties in processing L2 information (i.e. reading and listening). If it is the case that these skills rely on effective predictive or anticipatory mechanisms which one then relates to the stream of incoming speech (or the written text), then one is hampered by not having appropriate semantic networks established by means of which one can efficiently make appropriate comparisons.

In other work conducted by Meara (1984) he investigates 1) the ways speakers of different languages organize their mental lexica and 2) how learners of languages which use a non-Roman alphabet learn to visually process words in Roman script. In the first case Meara reports that native speakers of Spanish rely more heavily on syllables as a means of representing words than do English speakers. While he does not report on studies of learners, one can infer that when fundamental organizing principles of the lexicon differ, as in the case of English and Spanish, learning problems will result. In the second case, Meara considered Chinese speakers learning English. It appears that rather than reading words as wholes, they construct English words out of parts. This suggests that learning the lexicon involves much more than learning new meanings, ranges of meaning, collocations, and associations.

Bialystok (1987) reviews work comparing the development of the concept of 'word' by monolingual and bilingual children. She reports that bilingual children are better able to selectively attend to certain aspects of language than are monolingual children. For example, they can better separate strings into words even when the strings are actual sentences. For monolingual children, the inclusion of meaning (in the form of sentences as opposed to nonsense strings) interferes with their ability to focus on formal boundaries. Additionally, bilingual children are, for the most part, better able to separate the name of an object from the object itself, thereby showing greater understanding of the arbitrariness of word names. Bialystok argues that the metalinguistic abilities of the bilingual children serve them in general language development as well as in certain language skills such as reading.

In work by Levenston (1979) and Blum and Levenston (1979) lexical acquisition is seen as lexical simplification. They argue that in lexical acquisition there is avoidance of words which present phonological, grammatical or semantic difficulty. With specific regard to lexical simplification, they claim that learners prefer words which can be generalized to a wide range of contexts. Generalizing will take place even when it neutralizes collocational restrictions, register restrictions and when it falsifies relationships of hyponymy, synonymy and antonymy (Levenston, 1979). Examples are given below.

9. Collocations

Although we......them may compliments,
we received nothing but insults in return.

a. gave b. sent c. dropped d. paid

10. Register

In reply to your letter of November 17, 1977,
we regret to.....you that we are unable to meet your
request.

a. tell b. inform c. announce d. say

11. Antonymy

hatnu'a hafeministit makira behevdel sebein.....
la isa.

The feminist movement acknowledges the difference
between.....
and women.

In example 9, Hebrew learners of English do not use the appropriate collocation 'pay a compliment'. Rather, they select the more general term 'give'. In example 10, the formality of a business letter requires the use of the word 'inform', not the choice of learners. To understand example 11, it is necessary to point out that the word for woman in Hebrew isa serves all the functions of womanhood including the social and physical. However, the morphological counterpart for man is has a more restricted function. There are two other words: gever which in addition to meaning man also has the connotation of virility and baal which means husband. In example 11, 98% of the native speakers selected gever whereas only 55% of the learners opted for that word. 39% selected is. The authors argue that is and isa are learned as complementary forms. As a result, is has taken over the same semantic functions as isa, thus neutralizing the effect of the other antonyms of isa.

To conclude this section, I would like to extend the notion of interrelationships within the lexicon to relationships between the lexicon and other parts of the grammar. Ard and Gass (1987) question what has traditionally been viewed as syntactic acquisition. In particular, they argue that what appears to be syntactic acquisition can more appropriately be considered lexical or semantic acquisition. In their study there were two groups of learners, a Low proficiency and a High proficiency group. In general, they found an increasing importance of semantically based factors in lexical organization as a function of proficiency.

4. The role of the native language

I now turn to a discussion of the native language in the development of the second language lexicon. Native language influences in second language acquisition have been a topic of much debate in the literature of the past 10-15 years, with the pendulum swinging from the all-importance of the native language to the relative unimportance of the native language. In the past five to eight years investigators have been less concerned with the extremes of the pendulum swing and are more concerned with in-depth investigations into the ways in which the native language (or other languages known to the learner) affect second language development.

In work by Blum and Levenston (1979), they propose that learners avoid using words for which there are no direct translations in the native language. That is, if there is a semantic void in the native language lexicon, learners will not use the target language word which would fill that void. To test this hypothesis, they used a sentence completion task. While their data are skimpy in terms of the number of items from which to draw conclusions, their results suggest support for their proposal. The two sentences tested, 12 and 13, are given below.

12. yaldei olim bekitot regilot einam.......latsumat
halev sehem re'uyim la.

Translation: Immigrant children in regular classes
do not.....the attention they deserve.

13. hasefer mithaleq lehamisa peraqim. hapereq harison
.....bayehasim sebein medina ahat liskhenoteha.

Translation: The book is divided into five chapters.
The first chapter.....with relations
between a state and its neighbors.

In 80-88% of the cases Israelis respond with zakha 'to obtain, to receive, AS OF RIGHT' for sentence 12. For sentence 13 they respond with dan which combines the meanings of 'deal with, discuss'. Native speakers of English whose language has no direct equivalent of these words choose the expected word 11-13.5% of the time. Blum and Levenston do not provide an explanation as to why learners would avoid semantic voids. However, it may have to do with the nature of learning itself. In a recent paper (Gass, in press), I suggested that selective attention is a primary factor in determining what is learned. I also argued that prior knowledge is a motivating force behind selective attention. Therefore, because the native language has no semantic concept which is equivalent to the target language word, there is no foundation on which learners can build. Learners cannot easily focus on an undeveloped (or perhaps underdeveloped) concept. The existence of a concept thus serves as the starting point from which learners can figure out the correct semantic boundaries/parameters for a particular word.

As mentioned earlier, the concept of transfer has had a rocky history in recent years. Lado 'popularized' the notion in his 1957 book. His position is typified by the oft-quoted excerpt below:

> .....individuals tend to transfer the forms and meanings and the distribution of forms and meanings of their native language and culture to the foreign language and culture--both productively when attempting to speak the language and to act in the culture, and receptively when attempting to grasp and understand the language and the culture as practiced by natives (1957, p. 2)

This view has been challenged by many. With regard to the lexicon, Ard and Homburg (1983) propose that transfer be viewed not only as a direct imposition of native language forms (and meanings) in a second language situation. Rather, they propose a return to the original meaning of "facilitation of learning." Their work was based on a comparison of responses of native Spanish and native Arabic speakers to the vocabulary section of a standardized test of English as a second language. Clearly, one would expect that the greater similarity between the lexical items of English and Spanish would yield higher scores on the part of the Spanish speakers when compared with the Arabic speakers. In fact, this was the case. However, in addition, the response patterns of these two groups of learners differed even on those items in which there is no similarity between the native language and target language word. Ard and Homburg claim that these results represent transfer in the original sense of facilitation of learning. In other words, native language influences can be seen even in the absence of overt similarity.

Kellerman (1979, 1983) in his work on core meanings of lexical items has argued that there are constraints on language transfer which go well beyond mere similarity and dissimilarity of the native and target languages. These constraints involve the learner as an active participant in the learning process, one who makes "decisions" about what can and cannot be transferred. Kellerman suggests two interacting factors which are involved in language transfer. One is the learner's perception of the nature of the L2 and the other is the degree of markedness of an L1 structure. Language transfer, in Kellerman's view, is least likely when the perceived distance between the L1 and the L2 is great and when a given L1 structure is perceived as marked. These two factors interact so that a marked structure may be transferable if two languages are perceived as close; and conversely an unmarked structure may not be transferable if two languages are perceived as distant.

Kellerman refers to the perception of the distance between the L1 and the L2 as "psychotypology." A learner's psychotypology develops on the basis of many factors, not the least of which is actual linguistic typology. For example, an Italian speaker learning Spanish is likely to notice very early that there are numerous similarities between her/his native language and the target language. This will lead the learner to assume the potential success of forms or organizational principles from the native language when using the target language. On the other hand, a Japanese speaker learning French will be unable to find many correspondences between the native and target languages in most grammatical domains. Since the learner is unable, at least initially, to establish relationships between the two languages, s/he will not transfer from the native language. According to Kellerman, there are additional constraints on the transferability of a linguistic form based on the learner's organization of the native language. That is, there are parts of one's language which native speakers consider irregular, infrequent or semantically opaque. These items are 'marked' and are less transferable than frequent and regular forms. The former are language specific while the latter are language-neutral. Language-specific elements are those which a learner views as unique to his/her language, whereas language-neutral elements are those which the learner believes to be common to at least the native and the target language. It is a relative notion, since what is language-neutral with regard to a particular L2 may be language-specific with regard to a different L2. In other words, it is dependent on the learner's perception.

To relate this directly to issues of the lexicon, Kellerman experimented with various meanings of the word break as in 14-18 below:

14. She broke his heart.
15. The man broke his leg.
16. She broke the glass.
17. The tree broke his fall.
18. A game would break up the afternoon a bit.

He tested Dutch learners of English. Dutch has a similar word breken which can be used in many of the same senses as the English counterpart. Through testing of Dutch native speakers, Kellerman first identified what he calls 'core' meanings of the word 'break'. The core meanings are those meanings exemplified in 14-16, whereas sentences such as 17 and 18 have meanings which are further removed from the core. The core has to do with a variety of factors including frequency and perhaps abstractness in this case. No single factor however can account for differences. With regard to second language learning, he found that native Dutch speakers were most likely to think that the core meanings of 'break' could be used in English and that the more peripheral ones could not be used, despite the fact that they exist in both Dutch and English.

In related work Johnson and Sheldon (1986) investigated the use of false cognates between closely related languages. Their subjects were native speakers of Norwegian who had knowledge of Swedish. In their paper they argued that language distance and language similarity are not monolithic phenomena. The acquisition of cognates clearly includes notions of similarity, but additionally takes into account frequency, context, grammatical class and the nature of the semantic relatedness of the two words in question (cf. Ard & Homburg, (1983), for further discussion of phonological and semantic similarity).

5. Language use

In this section I will discuss issues related to language use. Due to the extensive literature on communication strategies, I will not include this topic in my discussion.

Olshtain (1987) investigated the development of word formation process by learners of Hebrew. She considered the use of new words from the perspective of production and interpretation. In the case of production tasks learners showed a progression toward target language patterns. For example, native speakers tended to provide innovative words 74% of the time versus 67% for advanced learners and 19% for intermediate learners. Additionally, when intermediate learners did innovate, the mechanisms they used differed from those of both advanced learners and native speakers. For example, intermediate learners used no instances of blending, whereas native speakers and advanced learners used 20% and 14% respectively. Blending refers to word combination as in 19 and 20 below.

19. English
    motel (from motor + hotel)

20. Hebrew
    migdalor (lighthouse)

    from 'migdal' (tower) + '?or' (light)

Similarly, intermediate learners used affixation as a word formation device 61% of the time, while only 35% of native speaker and advanced learner innovations contained affixation. However, on the interpretation task, a task which required learners to assign meaning to new words, neither group (advanced or intermediate) of learners performed like native speakers. On this task learners were asked to interpret words out of context. The inability to interpret words in a native-like fashion, even by those who produced words like native speakers, suggests that even advanced speakers are highly context bound in their use of the L2.

Blum-Kulka and Levenston (1987) emphasize the importance of the lexicon from a pragmatic point of view. Their study was concerned with the use of lexical and grammatical pragmatic indicators by learners of Hebrew and English. In general, learners of Hebrew differed from Hebrew native speakers along two dimensions: perspective and external modification. The first refers to differences in orientation (speaker versus hearer) as in:

21. Hearer oriented
    Can you do it?

22. Speaker oriented
    Can I borrow your notes?

23. Speaker and hearer inclusive
    Can we clean the kitchen?

24. Impersonal
    The kitchen needs cleaning.

External modification refers to the presence/absence of information about a particular speech act---information such as the underlined portion in 25.

25. I missed the class yesterday, could I borrow your
notes?

In the Hebrew as an L2 data, the authors noted the use of impersonal requests which could be linked to lexical simplification (see section 3). An example is provided from the Hebrew construction efsar (possible) + infinitive which is used as a means of request. This construction, however, is limited to verbs which express a speaker's perspective as in 26.

26. efsar leqabel et hamaxberet selax leqama
yamim?
Is it possible to have your notebook for a
few days?

Non-native speakers seem unaware of this constraint. As a result, they simplify it by generalizing the construction to include other perspectives as in 27.

27. efsar latet li et hamaxberet
Is it possible to give me the notebook?

As the authors point out, lexical simplification in instances such as these have pragmatic and interactional consequences.

A final paper dealing with the use of the lexicon is by Blum-Kulka and Olshtain (1986). Their focus is on pragmatic failure which they define as occurring "whenever two speakers fail to understand each other's intentions." An interesting finding is that learners tend to diverge from native speaker norms by lengthening the speech act by adding additional modifications as in 25 above and 28.

28. Hebrew as L2 (translated)
Please start cleaning the kitchen whenever you have
time,
and I prefer right away.

The development of target like use with regard to quantitatively appropriate speech has two dimensions: proficiency and length of residence in the target language community. As far as proficiency is concerned, there is a straight path from shorter utterances (due in part to limited knowledge of the language) to longer utterances (exceeding target language norms). On the other hand, with regard to length of stay (which may or may not correspond to proficiency level), there is a decreasing trend of utterance length. It is only after five to seven years of exposure that the quantitiative use of words approaches that of target language usage.

6. Conclusion

This paper has provided an overview of the L2 lexicon based on research conducted during the past decade. I have argued for the importance of the lexicon in terms of research on L2 knowledge in general. Unfortunately, neither L2 research nor L2 pedagogy has focussed on this area to any great extent. It is my hope that the research reviewed can stimulate our thinking about ways to implement research findings within the pedagogical arena.

Note

[1] Palmberg (1987) notes other terminology referred to in the literature, such as active real vocabulary versus passive real vocabulary. However, even with other terminology, there is still a binary distinction being made.

References

Ard, J., & Gass, S. (1987). Lexical constraints on syntactic acquisition. *Studies in Second Language Acquisition, 9,* 235-255.

Ard, J., & Homburg, T. (1983). Verification of language transfer. In S. Gass & L. Selinker (Eds.), *Language transfer in language learning*. Rowley, MA: Newbury House.

Beckwith, R., Rispoli, M., & Bloom, L. (1984). Child language and linguistic theory: In reponse to N. Hyams. *Journal of Child Language, 11,* 686-688.

Belasco, S. (1965). Nucleation and the audio-lingual approach. *Modern Language Journal, 47,* 482-491.

Bialystok, E. (1987). Words as things: Development of word concept by bilingual children. *Studies in Second Language Acquisition, 9,* 135-142.

Bialystok, E., & Sharwood Smith, M. (1985). Interlanguage is not a state of mind: An evaluation of the construct for second-language acquisition. *Applied Linguistics, 6,* 101-117.

Blum, S., & Levenston, E. (1979). Lexical simplification in second language acquisition. *Studies in Second Language Acquisition, 2,* 85-94.

Blum-Kulka, S., & Olshtain, E. (1986). Too many words: Length of utterance and pragmatic failure. *Studies in Second Language Acquisition, 8, 165-179.*

Blum-Kulka, S., & Levenston, E. (1987). Lexical-grammatical pragmatic indicators. *Studies in Second Language Acquisition, 9,* 157-172.

Gairns, R., & Redman, S. (1986). *Working with Words*. Cambridge: Cambridge University Press.

Gass, S. (in press). Integrating research areas: A framework for second language studies. *Applied Linguistics, 9, 1.*

Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning, 24,* 287-298.

Hakuta, K. (1976). A case study of a Japanese child learning English as a second language. *Language Learning, 26,* 321-352.

Hatch, E. (1983). *Psycholinguistics: A second language perspective*. Rowley, MA: Newbury House.

Johnson, T., & Sheldon, A. (1986). The interpretation of false cognates. Paper presented at AAAL annual meeting. New York.

Kellerman, E. (1979). Transfer and non-transfer: Where we are now. *Studies in Second Language Acquisition, 2,* 37-59.

Kellerman, E. (1983). Now you see it, now you don't. In S. Gass & L. Selinker (Eds.), *Language transfer in language learning*. Rowley, MA: Newbury House.

Lado, R. (1957). *Linguistics across cultures*. Ann Arbor: University of Michigan Press.

Levenston, E. (1979). Second language acquisition: Issues and problems. *Interlanguage Studies Bulletin, 4,* 147-160.

Meara, P. (1978). Learners' word associations in French. *Interlanguage Studies Bulletin, 3,* 192-211.

Meara, P. (1984). The study of lexis in interlanguage. In A. Davies, C. Criper & A. Howatt (Eds.), *Interlanguage*. Edinburgh: University of Edinburgh Press.

Ohshtain, E. (1987). The acquisition of new word formation processes in second language acquisition. *Studies in Second Language Acquisition, 9,* 223-234.

Palmberg, R. (1987). Patterns of vocabulary development in foreign-language learners.

Studies in Second Language Acquisition, 9, 203-222.

Pike, K. (1960). Nucleation. *Modern Language Journal, 44*, 291-295.

Politzer, R. (1978). Errors of English speakers of German as perceived and evaluated by German natives. *Modern Language Journal, 62*, 253-261.

Sharwood Smith, M. (1984). Discussion of Meara's The study of lexis in interlanguage. In A. Davies, C. Criper & A. Howatt (Eds.), *Interlanguage*. Edinburgh: University of Edinburgh Press.

Tarone, E. (1983). Some thoughts on the notion of 'communication strategy'. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication*. London: Longman.

Telchrow, J. M. (1982). A survey of receptive versus productive vocabulary. *Interlanguage Studies Bulletin, 6*, 3-33.

Varonis, E., & Gass, S. (1985). Non-native/non-native conversations: A model for negotiation of meaning. *Applied Linguistics, 6*, 71-90.

DISCUSSION OF PAPER OF SUSAN M. GASS "L-2 VOCABULARY ACQUISITION"

James R. Child

Dr. Gass in considering the attainment of second-language vocabulary from an acquisitional rather than pedagogic point of view raises some extremely interesting questions. Commencing with a discussion of the centrality of the lexicon in language use she continues in the following four sections to expand upon this theme as it relates to linguistic theory (2), connections within the lexicon (3), the role of the native language (4), and language use (5).

Finally, in her brief conclusion she points to the regrettable paucity of research findings, both general and pedagogical, on vocabulary acquisition in second-language study.

I will try to do as much justice to Gass' points as the scope of the paper permits. In so doing, however, I will take a point of departure somewhat different from hers, and, indeed, from that of many scholars who concentrate on either language acquisition or formal learning. Nonetheless my operational framework attempts to account for language use can accommodate both language norms as established by linguistic analysis and the natural processes involved in language production and comprehension. Before considering Gass's paper in detail I will outline my linguistic frame of reference with special emphasis on the role of lexicon.

Following the practice of the linguistic founding father, Ferdinand de Saussure, I begin with relationships: between sounds (letters) and "words"; "words" and the grammatical roles they play; grammar vis-à-vis semantic function (i.e., as grammar is used in the process of communicating); and semantic units and the effects they are intended to have in particular situations. I will refer to these four relationships respectively as phono-morphological (p-m); morpho-lexical (m-l); lexico-semantic (l-s) and semo-pragmatic (s-p). Each of these can be considered as a perspective from which language can be viewed either in terms of the relevant relationships between, say, form and function in general or particular forms or functions. It is rather like referring to water as H_2O on the molecular level, with the subsequent breakdown to H2 on the one hand and O on the other when atoms rather than molecules are of concern. Below I will attempt to establish the domains of the respective relationships, selecting as examples some of the usages Gass cites. Since her paper centers on the problem of acquiring vocabulary in a second language I will keep in mind at all points the pervasive structural differences between any two languages which, after all, are criterial for establishing the products of particular speech communities as languages in the first place.

In specifying the four relationships noted above I gave them in the order phono-morphological (p-m) through semo-pragmatic (s-p) although the "discovery" or identification of the linkages proceeds in the opposite direction, which is to say, analytically rather than synthetically. Practically, it is easier to go from the bottom up (although there are theoretical problems in that course) in treating language issues in general and vocabulary in particular.

The p-m bond, like the other three, can be identified in both native and second languages, written or spoken. Of particular interest in this relationship is the consonant-vowel syllabic structure underlying what for want of a better term I will call "words";

entities that from this p-m perspective may evoke mental associations but whose grammatical roles are not fully specified (or not specified at all). The issue here is usually one of spelling or pronunciation: A teacher may read words aloud from a list and ask the class to spell them correctly, or may give the group a list of target-language words or phrases and ask them to read the material aloud. Of course in the case of learning, or even acquiring, a second language the phonological characteristics of the native language may be obtrusive. Still, except in special cases such as the one noted above, or a true-to-life situation where the text is extremely limited and the context mainly physical, the participants in an exchange can usually make themselves understood from extended verbal context. To cite the exchange introduced by Gass early in her paper, José has a problem with the "urinals" and asks the befuddled teacher for help. She finally figures out that José is talking about "journals" and the difficulties he was having with a journal assignment. Had he merely seen the word "journal" somewhere and tried to vocalize it in asking what the word meant the problem most likely would not have been solved. That is why pedagogical materials need to be organized in as natural a way as possible, i.e., in simple textural form rather than in word lists, and why persons dealing with a new language system should try to acquire more than single items one at a time. One undoubted value of a discrete item approach is that it acquaints the learner with the various canonic shapes of a given language system and thus serves in the study of language typology or history. Unfortunately this has a limited utility for most students.

This naturally leads to the second perspective, the association of form with grammatical categorization. Labelled morpho-lexical (m-l), the issue is one of contextualization at least to the point of categorizing form-classes in the longest formal unit (according to Halliday and Hasan), the sentence. The m-l perspective is concerned with affixation, intonation and word order, or, in other words, the syntax of the sentence. Syntax underlies, of course, all cognitive and affective discourse. In being systemic rather than process-oriented it allows for the bizarre and the virtually meaningless as well as usual, expectable semantic structures. Thus, time-honored "school-book" sentences such as "The dish is on the table", or "John gave the book to Mary"; or peculiar ones such as "The son of the emir has five orange pencils"; or even the famous Chomskyan "Colorless green ideas sleep furiously" all have a formal validity equivalent to any text-oriented sentence. Each of these sentences contains a case-frame in which verbs and nouns (in particular) occupy slots specified by the syntactic rules of English. Obviously the sentences cited above become so many algebraic formulae, with the putative "vocabulary" entries substitutable by x number of others. Their raison d'être is to illustrate grammatical points of interest in the native language, or to test second-language knowledge of similar points.

Fortunately Gass avoids this perspective on vocabulary because she concentrates on how language is acquired rather than the way it is presented (at least much of the time) in school texts. People attempting to acquire a second language are more interested in communication than in algebra so that they ordinarily do not try to conjure up sentence-length paradigms.

Nonetheless even in the process of language acquisition, vocabulary as an m-l problem does arise. For whatever reason--and our knowledge of the mental processes productive of language is still scanty--formal/functional isomorphism is an important fact of language behavior especially where two languages are simultaneously concerned. Gass cites the example of erroneous but recoverable ordering in "Can you tell me where is the train station?" in which the questioner wrongly uses the direct question order for the indirect, very likely governed by the use of identical syntactic ordering in his or her native language. Vocabulary misuse is more dramatic and frequently more difficult to recover. In Gass's example "What tragedies (for strategies) do I have to follow?" the speaker has, perhaps inadvertently, produced a legitimate morphone string (tragedies) and put it in a syntactically correct slot, causing at least a temporary breakdown in communication since the professor to whom the question was addressed "lapsed into silence." Had the questioner come up with the nonsense form stragedies (a phono-morphological slip of the tongue) the professor might have recovered the information immediately.

Before passing to the lexico-semantic perspective of critical import to the question of vocabulary mastery, I will comment briefly on a study cited by Gass of native speakers learning French (Meara, 1978). Meara claims that mental lexica are organized semantically so that in word associations "a native speaker is likely to come up with words...similar in meaning to a stimulus word," and if the person is an adult, he or she "is more likely to come up with paradigmatic associations (than) syntagmatic ones." An example of each type is cited by the author, the English stimulus/response table provided as an introduction to the French:

stimulus	response
a) woman	(paradigmatic)
b) brush teeth	(syntagmatic)

Zeroing in on a word at a time (usual in word association exercises) on the surface of it suggests an m-l approach to vocabulary but has lexico-semantic (1-s) implications as well: "man is a human, " "woman is a human," but "man is of the male sex," "woman is of the female sex," etc. Silly as all of this sounds it may well be that an 1-s "definition" in which x is associated with y undergirds the adult English speaker's behavior. In the case of "brush" and "teeth" the native speaker (presumably a child) focuses on an instance or example of some purpose brushes serve. The mental "storage problem" may be different for each group but an l-s "projection" is certainly plausible.

Interestingly, as the author points out, the situation with learners is different. She employs Meara's table to highlight their problems (in this case with French) as follows:

From Meara (1978)

Stimulus	Native response	Learner responses
profond "deep"	puits "well"	plafond "ceiling" prendre "take" professeur "professor"

Here, quite obviously, the learner's "grasp" of the French vocabulary item is so tenuous that he or she focuses not on the grammatical category, much less the meaning of the word, but on its canonic form as perceived by a speaker of English. Segments of the French formal string trigger responses which are also of a formal nature but have no discernible semantic association with the stimulus. In other words, "vocabulary" in this context is a question of pure form (a p-m issue). This may tell us that the morpho-lexical approach to teaching, with one-to-one relationships between form and function, may not only fail to develop for the learner a communications strategy but actually inhibit his or her progress. And even the acquirer may suffer from a mindset resulting from earlier experience with paradigmatic learning of some other language or "grammar instruction" in his or her own (I expand on this point later).

As will be clear from above, the problem of vocabulary acquisition extends beyond phonological composition and grammatical role to the roles of words (or more properly, word groups) in expressing feelings, opinions or facts as textually relevant. As Gass observes in her short section on acquisition and linguistic theory acquisition research (whether first or second language) and a theory of language must be responsive to one another, a proposition critical for semantic theory as it applies to contextualized lexicon. The lexico-semantic (1-s) perspective requires a sophisticated understanding of the relationship of formal segments to mental constructs.

In addressing this point Gass quotes Pike (cited in Belasco) on "nucleation," a physical phenomenon whereby atoms or molecules cluster into a structural pattern, eventually forming a crystal: "It is difficult to get the first molecules to clump together...yet once nucleation has begun, growth may proceed with great rapidity." This seems to me a highly apposite analogy especially for persons who have some knack for "picking up" a language and who are unencumbered by the discrete-item approach to grammar and especially vocabulary. For those formally learning a second language the comparison will hold only if the instructional materials are designed to promote "nucleation," i.e., when they duly account for the varying degrees of context sensitivity required for communicating a wide array of facts, opinions and feelings. The usual "words" to be mastered lesson by lesson are often without obvious interconnection, as Gass points out. Worse yet, they generally appear in sets only in the sublists on "idioms"; otherwise they are confined to dictionary-entry listings (verb in infinitive, noun in subject form, if relevant, etc).

For example, suppose we have a lesson in an English text which includes the word "take" as a vocabulary item. Other than the idiom "to take one's time" given at the bottom of the list there may be no indication that "take" can mean anything but "remove from a location" or "accept from a person." Thus, the student may conclude that "take a seat" simply means "move a chair from its place," or that "He took the money" means "He accepted it from someone," without the possible alternative "He stole the money." He may also think that it is as permissible to say "Pursue your nose" as "Follow your nose" in the absence of information on verb-noun collocation.

In Section 4 Gass covers similar ground in quoting E. Kellerman's studies (1979, 1983) on psychotypology, definable as the perceived distance between the native language and the language being acquired. The latter identified "core" and peripheral meanings for the Dutch verb "breken" (cognate with the English verb "break"). He then selected some examples of core meanings of the English "break," such as "She broke the glass" and of peripheral meanings such as "The tree broke his fall." Experimenting with a group of Dutch native speakers learning English, he found that the latter tended to believe that the core meanings transferred from Dutch to English, but the peripheral ones did not.

In my examples I had in mind the formal learning situation whereas Gass and Kellerman were apparently concerned with acquisition (it is not clear from Gass's paper whether Kellerman was carrying out his experiments in the classroom or outside it). It is interesting though that both points of departure concerning the role of vocabulary seem to be grounded in the m-l perspective, in which case it could be inferred that the text book bias toward the discrete item approach merely reflects how second-language learners in general naturally behave.

In this connection Gass cites the study of Meara (1984) in which he reported that "native speakers of Spanish rely more heavily on syllables as a means of representing words than do English speakers." Obviously such phonological representation will not do violence to the grammatical structure for native speakers of Spanish: An accurate transcript of a spoken text will not exactly correspond to the written norms of the language in regard to such elements as juncture, for example, but literate speakers of the language process both written and spoken texts without difficulty. However, in transferring their language habits to English they may violate syllabification and juncture rules (a p-m problem), thus complicating the relationship of the morpho-lexical to the lexico-semantic even further. The native speaker of English conversing with the Spanish speaker may have to make a tremendous effort to establish the message first through the grammatical (m-l), then the phonological (p-m) system (or even in reverse order). Also according to Meara, when speakers of Chinese transfer their phonological habits (doubtless reinforced among the literate by the writing system) to English they "construct English words out of parts", thus producing a kind of spoken rebus. As Gass points out, knowing vocabulary entails knowing "much more than learning new meanings, ranges of meaning, collocations and associations". In my terms it means calibrating the l-s to the m-l to the p-m, a formidable task not only for the preparation of instructional materials but also for anticipating the language acquirer's need for reinforcement and correction.

Much more could be said about the l-s perspective but the final test of the communicative process depends on the ability of the language learner not only to produce correct, idiomatic communications (i.e., texts) but to understand what effect these will have on listeners and to properly interpret the responses to them. Here we are in the area of pragmatics, a complex domain of language do's and don'ts. Language use under this rubric becomes a matter of appropriateness and vocabulary, a question of effect: the use of tabooed, profane or obscene words or phrases to cause shock; the use of suitable phrases to allay fear or suspicion; low-key words in reporting unpleasant or otherwise negative facts and events, and the like. This perspective I refer to as the "semo-pragmatic" (s-p), not the "pragmatic" alone, since the occasion for the particular effect(s) is the cognitive or affective content embodied in the verbal structure.

To this point I have not specifically introduced the question of the effect of one's native cultural presuppositions in learning a second language. In the case of the p-m and probably the formally oriented p-m and m-l perspectives the problem is negligible. The l-s view of language obviously does involve cultural values and biases but those phenomena are better illustrated in concrete examples embedded in language transactions at particular times and places, a virtual definition of the semo-pragmatic.

An example of the s-p perspective may be a sentence uttered by a Spanish speaker which Gass cites early in her paper: "I feel sorry for people who live in the suburbs." In that context she was discussing the kinds of language error (phonological, grammatical and so forth) which can be recovered from context and which cannot. There is no indication in the sentence as transcribed of any phonological, or grammatical problem so the issue seems to be lexico-semantic or, better, semo-pragmatic, since the statement was ascribed to a particular conversation. The expression may indeed have been a reflection of the sympathy of the speaker for the benighted condition of suburbanites, or merely an attempt to touch off verbal fireworks. The precise motivation is certainly irrecoverable but that probably has more to do with psychological motivation than error in communication. The example comes to mind of the Greek general who sought advice from the oracle at Delphi regarding his prospects in the next-day battle with a formidable army. Told that at the end of the day a great army would be destroyed, he confidently (and wrongly) assumed that it was the enemy's. Oracles, like ordinary mortals, have private purposes not necessarily associated with linguistic error.

In addition to the cultural taboos and oracular ambivalence (itself a specific cultural phenomenon) the s-p perspective involves more mundane differences between languages, as Gass points out in some examples she cites from Hebrew. To mention one of these, a student asks another "Is it possible to have your notebook for a few days?" (literal rending from the Hebrew), i.e., "Can you lend me your notebook...?" In Hebrew the speaker's perspective is stressed (at least in this kind of construction) while in English the point of view is the hearer's.

Dozens of such examples could be cited between any two languages, but an exhaustive s-p overview will require a tremendous amount of research and is no doubt a long way off.

Nonetheless, Gass has done an excellent job in identifying an impressive cross-section of errors by language acquirers and (on occasion) learners. My hope is that I have succeeded to some degree in offering an overview of the problem in the context of the phono-morphological, morpho-lexical, lexico-semantic and, finally, semo-pragmatic perspectives, and that further research will lead to refinement in the approach to lexical acquisition.

References

Bach, K., & Harnish, R. M. (1982). *Linguistic communication and speech acts.* Cambridge, MA: MIT Press.

Belasco, S. (1965). Nucleation and the audio-lingual approach. *Modern Language Journal, 47*, 482-491.

Hjelmslev, L. (1953). *Prolegomena to a theory of language.* (F. J. Whitfield, Trans.). Bloomington, IN: Indiana University Press. (Original work published 1943)

THE ROLE OF VOCABULARY IN THE DEVELOPMENT OF COMMUNICATIVE ABILITY: SOME COMMENTS[1]

Sharon Lapkin and Susanne Carroll
Modern Language Centre
Ontario Institute for Studies in Education

We want to take up Sue Gass's challenge to reflect on an appropriate research agenda for examining the development of vocabulary as a central component of communicative ability in a specific pedagogical context -- that of French immersion education. Although Canada is officially bilingual, this fact has little day-to-day impact on the average anglophone student of French as a second language whose sole contact with French tends to be confined to the classroom. Student surveys administered in the context of large-scale evaluations of French immersion programs have included items designed to elicit information on the amount of out-of-school contact with the second language -- whether through relatively passive activities such as watching television or through active efforts to meet and interact with francophones in the community. Even where there are opportunities in the immediate environment for such interaction -- the survey data suggest that immersion students do not seek them out (Lapkin & Swain, 1984, 1985).

For the majority of students, then, French language input is largely confined to the classroom, thus providing a situation where input is tightly controlled. Moreover, the fact that most instructional time is devoted to teaching through the medium of French means that the students' second language skills reach relatively high levels of development. What an ideal laboratory for examining the development of lexical knowledge and skill in a second language! Garrett (this volume) has stated that "we must have some idea of what learners are actually doing with language before we can study the effect of manipulating pedagogical variables". In this short paper we will provide some information about (1) what immersion students can do in French, with specific reference to lexis, (2) what teachers are doing by way of vocabulary instruction in some immersion classrooms, and (3) a possible research plan for investigating vocabulary learning in this context, given what we know about the properties of words and the organization of the lexicon.

1. Immersion students' lexical performance

Data reported here come from early immersion classrooms where students have had all their early instruction (kindergarten through grade 3 in most cases) entirely in French, followed by at least half of their school program delivered in French until the end of grade 9. After grade 9, the number of subjects taken through the medium of French varies from two to four (one quarter to one half of their program).

In recent years we have been collecting second language data from immersion students at the secondary school level through the use of a testing unit designed to resemble a teaching unit -- materials built on students' interests, consisting of thematically linked oral and written tasks intended to elicit the best performance possible (Green & Lapkin, 1984; Swain, 1984).

One of the written tasks is a composition which is assessed for lexis, along with other grammatical and discourse features. The lexical measures involve counting the number of verb types to obtain a sense of lexical range and determining how many of these types can be regarded as 'sophisticated' by assigning a value of "1" to each verb which does not appear among those listed as most frequent in Le français fondamental (Gougenheim et al., 1964).

Baseline data on the same task have been collected from comparison classes of monolingual francophone students from Quebec. Quantitative comparisons presented in Table 1 suggest that immersion students become more nativelike on these particular lexical measures as they progress through the secondary grades.

Aside from these quantitative results, which are confined to verbs and have not encompassed other lexical categories, there are at least two striking gaps in the type of information we have on the lexical proficiency of immersion students. The first relates to their receptive proficiency -- how well do immersion students understand spoken or written lexical items (word recognition in isolation as well as word recognition in context)? Evaluation studies (e.g. Swain & Lapkin, 1982) have suggested that many immersion students achieve nativelike listening and reading comprehension skills by grade 8, but such statements are based on global measures which emphasize deriving the main message from extended written text or oral discourse. We know of no studies focussing on the word recognition skills of immersion students or on their ability to identify new lexical items based on what they already know about French. Secondly, returning to the productive domain, what qualitative information have we on the vocabulary of immersion students? A recent study by Harley et al. (1987) compared verb vocabulary in grade 6 immersion student writing samples (two compositions and three request letters per student) with that of native French-speaking peers. There were four main findings arising from hypotheses based on the work of Levenston (1979) and Adjemian (1983) among others:

1. Immersion students used a higher proportion of high-coverage verbs (i.e., words with less specific denotations such as faire) than native speakers.

2. Immersion students used a lower proportion of syntactically and morphologically complex verbs such as pronominal and derived verb types than native speakers.

3. Semantic and/or syntactic incongruence with English L1 plays a major role in immersion students' lexical and grammatical errors, as well as in their non-use of some French verbs.

4. Lexical similarities between English and French, especially cognates, help to increase the French lexical resources of immersion students.

This type of qualitative analysis has not been conducted at higher grade levels, where, from a quantitative perspective at least, it would appear that immersion students may have caught up to their francophone peers.

Table 1

Number of Verb Types and Sophisticated Verbs

Used by Immersion and Francophone[a] Students Across Grades

Grade	Verb Types						Sophisticated Verbs					
	Immersion			Francophone			Immersion			Francophone		
	$\bar{X}$	SD	N	$\bar{X}$	SD	N	$\bar{X}$	SD	N	$\bar{X}$	SD	N
9 (Carleton)	15.60[b]	5.74	15				1.93[d]	1.67	15			
9 (Toronto)	13.43[c]	2.82	7	20.58	5.01	19	1.14[d]	1.07	7	5.47	2.89	19
9 (New Brunswick)	13.93[c]	6.13	70				1.54[d]	2.52	70			
10 (Carleton)	16.15	4.88	13	18.33	7.63	6	2.46	1.61	13	3.67	4.89	6
10 (Toronto)	18.29	3.20	7				2.71	1.50	7			
12 (Carleton)	18.92	4.86	24	16.56	6.85	100	3.92	2.38	24	3.81	2.68	100

a The grade 9 francophone group is of similar SES as the immersion groups. This also holds at grade 10; however, the francophones are bilingual (from a French-language high school in Ottawa). The comparison group for grade 12 immersion students are in their graduating year, which in Quebec, is the equivalent of grade 11. Also, these students are not as academically oriented as the immersion groups.

b Score is significantly lower than that of francophone comparison group at $p \leq .05$

c Score is significantly lower than that of francophone comparison group at $p \leq .01$

d Score is significantly lower than that of francophone comparison group at $p \leq .001$

2. Lexical instruction in immersion classes

As virtually the sole source of spoken input, the teacher talk and instructional techniques focussing on vocabulary are of overriding importance in studying the developing lexicon of immersion students. Carroll (1987) has recently analysed transcripts from ten grade 6 immersion classes (along with some curriculum materials in use in those classes) with a view to relating what linguists know about the lexicon to instructional techniques which could facilitate the development of lexical proficiency in French. She derives the following (non-exhaustive) descriptive categorization of instructional activities from her data analyses:

planned/unplanned
systematic/haphazard
written input/oral input
L1 prior knowledge/L2 prior knowledge
direct control of item choice/indirect control interpretation/formal focus

The planned/unplanned dimension identifies the extent to which vocabulary instruction is seen as a purposeful activity, occupying a distinct place alongside other types of language-related instruction. Planned vocabulary instruction would involve decisions about what aspects of lexical information should be taught (e.g. derivational morphology, semantic networks) and how. Unplanned instruction allows for spontaneous teaching of lexis in response to learners' expressed or intuited communicative needs.

A related dimension is that of systematic versus haphazard instruction. Even if vocabulary is acknowledged to be an important part of the language curriculum, instruction can be focussed and systematic (e.g. teaching rule-governed relationships such as the formation of French adverbs by adding a suffix, -ment, to the feminine form of the adjective) or occasional and unsystematic.

Written input is often associated with more formal registers than spoken input. Similarly, there may be more overt redundance in spoken language than in written where the need to be explicit is greater in the absence of non-verbal aids to communication. (This has obvious implications for the choice of lexical items.) Although these distinctions are not absolute, the range, distribution and relative frequency of lexical items will clearly be affected by one's choice of medium. Hence the need for exposing immersion students (for whom literacy in the second language is as important as the development of oral communicative ability) to lexical use in both modes.

One of Harley et al.'s findings cited above suggests that knowledge of English cognates may facilitate the learning of certain items of French vocabulary. This is one example of the possible uses of the L1 in helping the second language acquisition. Carroll points out that the similarities in the graphic representation of English and French words appear greater than resemblances in pronunciation. Thus crosslinguistic pedagogical techniques might be better explored through reading and writing activities where words may be less familiar, while intralinguistic phenomena such as masculine-feminine alternation might best be studied through oral practice.

The direct/indirect distinction as it applies to the control of item selection for vocabulary instruction relates to who decides what specific words will be taught, at what stage of learning, and in conjunction with what other features of language. Lexis undoubtedly influences what reading materials the teacher will choose for her/his students.

In communicative classrooms, the perceived communication needs of the learners may determine what lexical items are used, and the natural tendency of more competent speakers to accommodate to their less competent interlocutors may mean that oral language is not as lexically rich and varied in a situation which may be characterized by indirect control of lexical choice.

Finally, the instructional focus may be on the meaning or interpretation of a specific word, or on formal features of that word (e.g. spelling, pronunciation).

Time does not permit an in-depth consideration of how these activities could facilitate L2 vocabulary acquisition (this is discussed in considerable detail in Carroll, 1987); but let us consider one example where the descriptors might be 'planned', 'systematic', 'written input', 'prior knowledge of L2', and 'indirect control', but where the primary focus is clearly on formal features of a lexical item.

In the following exchange, the teacher has planned a lesson on the vocabulary of the passage to be read, and begins with a question about the meaning of a lexical item:

(a) T: un "bottillon de paille" Qu'est-ce ça veut dire "un bottillon de paille"? D'où est-ce que ça vient ce mot-là "bottillon" Regardez-le un peu:
S1: bouteille
S2: beau
S1: bouteille
T: "botte" "botte" Je prends une botte, vous savez, une botte de paille
S3: C'est un mor/c'est comme de la paille qui est mis dans un morceau
T: Oui Vous vous rappelez le film/pas le film!/le ballet qu'on a vu ensemble qui était "La Fille Malgardée"? Vous vous rappelez les paysans qui sont venus avec des bottes de paille sur la tête/euh sur l'épaule?
S4: Et il y a un/
T: Quelque chose qui les attache ensemble
S5: Une corde
T: Corde Bon! On arrête les/les commentaires (irrelevant text) mais alors/Pardon! le mot "bottillon" est plus petit que botte Comme vous vous rappelez vous avez fait un exercice "livre: livret", "porte: Portière" euh
S1: maison maisonette
T: "maison: maisonette"
S3: ruisseau
T: "ruisseau: ruisselet"
S3: fille fillette
S4: garçon garçonnet
T: "nappe: napperon" Alors le mot "bottillon: c'est un petit/une petite botte de paille

To get at the meaning of the diminutive bottillon, the teacher directs students' attention first to the meaning of botte and then turns to the morphological distinction at issue. The formal distinction between botte and bottillon is addressed by recalling a previous exercise on the formation of the diminutive. Only once the analogy is established does the teacher provide a definition of the word.

Consider a second example from the same lesson:

(b) T: Qu'est-ce ça veut dire
"une pincée"?
S1: un tout petit
T: Un tout petit peu! (irrelevant text) Faites-moi/ Comment est-ce que vous dites une pincée de sel, oui? (student mimes action) Voilà! Tu prends comme ça (gesture) avec les/les doigts
S2: un 'pinch' (long exchange about significance of throwing salt over one's shoulder)
T: Ça (gesture) c'est une pincée de sel Comment est-ce qu'on appelle ça? (gesture)
S3: une poignée
T: Poignée Bravo! Une poignée, n'est-ce pas? Une poignée de eum de noisettes, une poignée de riz, n'est-ce pas? et une pincée de sel. Très peu, c'est-à-dire on/ comme une pince

These exchanges are informative and spontaneous in that they reveal meaning in a systematic fashion and clearly indicate the function of the suffix which is the focus of attention. Furthermore, the teacher provides examples of uses of the words which are natural and spontaneous. They refer to and utilize the experiences of the students. The only minor improvement one might suggest to reinforce the structural comparisons would be to make explicit the relationship poing:poignée and pince:pincée as well as to mention other words illustrating the relationship (bouche:bouchée, cuillère:cuillerée, etc.). The idea would be to explicitly draw the learner's attention to the rule-governed nature of this relationship. Finally, these exercises illustrate the functional identification of the morphemes since they are used to focus on the expression of a given meaning (rather than just on the form itself). The diminutive is actually a morphological category rather than a semantic type but it corresponds reasonably well to a concept, namely a small (denotative) or cute (connotative) thing. So exercises which teach that -ette, -et, -on all express this concept are not only teaching form, but also function or meaning.

3. Research perspectives

The two examples given above illustrate good second language teaching techniques because although they are embedded in a fairly traditional explication de texte type of reading comprehension exercise, the students' attention is focussed explicitly and systematically on particular morphological generalizations that instantiate a specific semantic notion. Thus the text is being exploited for several purposes at once, but each is signalled as discrete, and the learner is permitted to focus on one thing at a time. The inventory of linguistic aspects that can be addressed through such an exercise includes phonology, graphology (sound-symbol correspondence), morphology, syntax, semantics, and discourse as well as sociolinguistic distinctions. Had the teacher chosen to do so, she might have focussed on some of these other dimensions. In text (a) *garçon:garçonnet* example serves as an example of a phonological rule whereby the nasalized vowel is denasalized under particular conditions. Some morphological properties of words such as *bottillon* or *poignée* were the focus of the lesson transcribed in (a) and (b) and have been discussed above.

The word *botte* has other meanings also, the most common being 'boot'. The possible exploration of polysemous meanings is an example of the semantic exploitation of the vocabulary item under study, and it can lead to a consideration of syntactic properties of its associated verb *botter* ('to shoe') or sociolinguistic nuances as in *ça me botte* ('suits me!'). Similarly, the use of diminutives as expressions of tenderness or emotion could have been examined as well as alternative strategies such as the use of *petit (ma petite)* or reduplication of epithets (*Il est mignon mignon mignon*.)

The possibilities are endless -- and to tackle all of these networks at once would almost certainly be a mistake. But to take all of these relationships among linguistic systems into account in selecting vocabulary for detailed attention in syllabus design and in planning related instructional activities is merely to pave the way for more effective learning of vocabulary and thereby of a broad range of linguistic knowledge. Moreover it provides the basis for studying the relationship between what is done in the classroom and the knowledge that students acquire.

At a minimum we have identified three major research thrusts we intend to pursue in the area of immersion acquisition of French lexis in the immediate future. The first involves obtaining data on the receptive L2 lexicons of immersion students to complement the information we have of a 'product' nature. The second is to develop more detailed descriptions of that product by conducting qualitative analyses of oral as well as written language data already on hand. The third and most important line of inquiry is also the most complex and involves manipulating syllabus content and instructional variables so that the relative effects of different pedagogical strategies can be assessed.[2]

In conclusion, we would like to make some comments about the role of vocabulary in the development of communicative language proficiency which includes grammatical, discoursal and sociolinguistic components. If sentences are indeed "projected" from the properties of the lexical items that they contain, as Chomsky (1981) has claimed, then a considerable amount of grammatical knowledge will consist of knowledge of the morphosyntactic, argument, and theta properties of words. If discourse competence involves minimally the ability to produce coherent and cohesive text (cf. Halliday & Hasan, 1976), then these will be established in part in terms of the relations defined within and across lexical sets in given semantic fields. Recognizing and exploiting the ways in which lexical items can define register, style and dialect must be part and parcel of sociolinguistic competence. Thus vocabulary as an 'entity' cuts across these components. We feel it is both appropriate and useful for future research to flesh out these components through the study of vocabulary in the developing grammars and interlanguage systems of the L2 learner.

Notes

[1] We wish to thank Doug Hart, Laurette Lévy and Razika Sanaoui for their comments on an earlier draft of this paper.

[2] For a recent experiment along these lines in the area of teaching past tense distinctions in French, see Harley, 1987.

References

Adjemian, C. (1983). The transferability of lexical properties. In S. Gass & L. Selinker (Eds.), *Language transfer in language learning*. Rowley, MA: Newbury House.

Carroll, S. (1987). Vocabulary instruction in immersion classes. In *Final report on the development of bilingual proficiency project* [Mimeo]. Toronto: Ontario Institute for Studies in Education.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Gougenheim, G., Michéa, R., Rivenc, P., & Sauvageot, A. (1964). *L'élaboration du français fondamental, ler degré*. Philadelphia: Chilton.

Green, D. Z., & Lapkin, S. (1984). Communicative language test development: Where we are and how we got there. *English Language Teaching Documents, 119*, 129-148.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Harley, B. (1987). Functional grammar in French immersion: A classroom experiment. In *Final report on the development of bilingual proficiency project* [Mimeo]. Toronto: Ontario Institute for Studies in Education.

Harley, B., King, M. L., & Burtis, J. (1987). Perspectives on lexical proficiency in a second language. In *Final report on the development of bilingual proficiency project* [Mimeo]. Toronto: Ontario Institute for Studies in Education.

Lapkin, S., & Swain, M. (1984). *Second language maintenance at the secondary school level: Final report to the Carleton Board of Education* [Mimeo]. Toronto: Ontario Institute for Studies in Education.

Lapkin, S., & Swain, M. (1985). Research update: The New Brunswick study. In W. R. McGillivray (Ed.), *More French, s'il vous plaît*. Ottawa: Canadian Parents for French.

Levenston, E.A. (1979). Second language lexical acquisition: Issues and problems. *Interlanguage Studies Bulletin, 4*, 147-160.

Swain, M. (1984). Large-scale communicative language testing: A case study. In S. Savignon & M. Berns (Eds.), *Initiatives in communicative language teaching*. Reading, MA: Addison-Wesley.

Swain, M., & Lapkin, S. (1982). *Evaluating bilingual education: A Canadian case study.* Clevedon, Avon: Multilingual Matters.

Summary of the discussion session on THE ROLE OF VOCABULARY IN THE DEVELOPMENT OF COMMUNICATIVE ABILITY

Laurel McLain
Indiana University

Discussion began with a response to S. Lapkin's comments on the role of vocabulary in developing communicative ability in French immersion programs. It was suggested that a potential advantage of immersion programs is that one could observe cultural as well as lexical concept development. Otherwise all that Lapkin presented could be carried out in a traditional language classroom. Lapkin and Carroll responded that even in an immersion program, exposure to target language culture was for the most part confined to what the teacher could bring to the students' awareness. One of the problems in teaching traditional academic subjects through the medium of French is the lack of adequate materials which match both the cognitive and language levels of the students.

The discussion then turned from curricular issues to those related to testing. How could the information derived from immersion research help in testing? How and to what extent should the guidelines specify the learners' lexical skills as compared, for example, with grammatical skills? It was pointed out that vocabulary is far more dependent on content or subject matter than is grammar, and it was felt that content neither is nor should be built into the various levels of the guidelines. Vocabulary depends on "purpose," which is more narrowly defined in the case of the government than in academia.

Lapkin then responded that language-specific guidelines could describe the word formation processes of learners based on further research of developmental patterns. She referred to Olshtain's study of the extent to which advanced learners of Hebrew could approximate native-like word formation processes. For example, advanced learners can use affixes in sophisticated ways, but fail to behave in native-like ways when it comes to productive rules involving word internal blends. Morphological information such as this could be inserted into the levels of the guidelines according to a researched hierarchy of difficulty.

Another central topic of this discussion was that of distinguishing between process and product. It was suggested that this distinction was crucial to teaching, and that although testers needed to be aware of process, they were forced to look at the product. The relation of process to product which occurs in program evaluation can be seen on a micro-level within an individual learner. In evaluating a program, however, it is easier to relate process to product. One cannot measure how language is acquired by an individual, but can only measure a "slice of life."

It was further suggested that the instructional variables Lapkin and Carroll presented in their paper could provide a way of characterizing test methods. For example, one could examine what a test would be like if it were unplanned rather than planned, haphazard rather than systematic, and indirectly controlled by the testee rather than completely controlled by the tester. Another participant mentioned a problem that test designers encounter: They can not give an ultimate reason for the exact items they select, and thus should not pretend to control this 100 percent.

In contrast, another participant maintained that the skills statements were aimed at describing process. This fact is often obscured by the numerous examples needed to adequately exemplify the level, and as a result, the descriptions risk being seen by some people as laundry lists to be covered. He continued that when trying to determine whether a testee is a 2+ or a 3, he regularly makes appeal to the learner's understanding of new material (e.g., from compounding rules in Germanic languages) in order to see whether or not processing is taking place. He noted that he was pleased to see, in the various symposia on proficiency testing, the emergence of more abstract terms which could ultimately be used to describe language processing. He recommended that instead of shelving the guidelines, as had been suggested at an earlier session, a third set should be written for the people at this symposium who wanted to know how to relate processing to the test. This would be a set of guidelines for internal use only, designed to give direction to research.

The response came that we know too little about processing to revise the guidelines accordingly. This revision is a longterm goal which does not answer the urgent question of how we are to test right now.

A final topic addressed at this session was the interdependence of learning, pedagogy, testing and the proficiency guidelines. These are not completely unrelated areas, but the degree of overlap remains unclear. It was suggested that perhaps three kinds of guidelines were needed: teacher, tester and user guidelines, the latter of which would not include any technical terminology, but would inform the user of what a person at a certain level could do with language in a prototypical real life situation.

It was pointed out that in the Netherlands there are two separate institutes for curriculum and test development, but that these work in close connection. A new curriculum plan can not effect change in education on its own, but must be accompanied by the control and beneficial "backwash" of a test. Nonetheless, these are two separate disciplines to be worked at in different surroundings.

Another participant added that although she agreed that there was disparity in theory between acquisition processes and testing, in the end, if the test was not about how acquisition occurred, what was it about? There was general disagreement at this point and the issue was left open.

There followed a suggestion that a dilemma had arisen from the fact that the guidelines were serving two separate testing purposes, one formal, and another, informal. In some cases, such as in the government, formal tests are required which are free from teaching situations. Informal testing is more directly tied to pedagogy and success in learning a certain language. The problem comes from trying to apply the same criteria to both cases.

In summary, this discussion began with a consideration of the role of teaching vocabulary and word formation processes in immersion programs in Canada. Information from immersion research was then applied to proficiency testing, and it was suggested that with more research, morphological information could be slotted into appropriate levels of language-specific guidelines. The discussion then turned to the distinction between process and product, and to whether or not it was desirable and possible to relate processing to testing. Finally, the interdependence of learning, pedagogy, testing and the proficiency guidelines was addressed.

ISSUES IN ESTABLISHING AND MAINTAINING A LANGUAGE PROFICIENCY REQUIREMENT

Barbara F. Freed
University of Pennsylvania

Within the last six years several foreign language departments at the University of Pennsylvania have taken the unprecedented step of redefining a traditional foreign language requirement as a proficiency-based requirement. In the process of making this transition numerous issues, philosophical as well as practical and administrative, were encountered. This paper provides an historical synopsis of Pennsylvania's decision to redefine its language requirement and a general description of the current proficiency-based requirement. Within the context of this discussion, attention focuses on numerous issues confronted in establishing and maintaining a proficiency-based foreign language requirement as well as on the improved language learning environment that has resulted from these changes.

Within the last six years several foreign language departments at the University of Pennsylvania have taken the unprecedented step of redefining a traditional foreign language requirement as a proficiency-based requirement. In the process of making this transition numerous issues, philosophical as well as practical and administrative, were encountered. This paper provides an historical synopsis of Pennsylvania's decision to redefine its language requirement and a general description of the current proficiency-based requirement as adopted by several foreign language departments. Within the context of this discussion, attention focuses on numerous issues confronted in establishing and maintaining a proficiency-based foreign language requirement.

Almost since its founding some one hundred and fifty years ago, Pennsylvania has maintained a foreign language requirement. Throughout this period the languages, of course, varied and the number of required semester hours changed. Nonetheless, the requirement itself changed very little. It was, in principle, a time requirement, known somewhat euphemistically by some as a "seat" requirement. That is, if students sat in their seats long enough, even with minimally passing grades they would eventually satisfy the language requirement. Beginning in 1979, Pennsylvania's Language Advisory Committee and members of several language departments began to examine closely the effectiveness of this requirement with the hope of redefining it in more current, efficient and meaningful terms. In addition to concern about the requirement itself, there was also concern about the fact that the foreign language requirement at Pennsylvania, as elsewhere, had not been (to say the least) one of the most popular aspects of the undergraduate curriculum. It was further believed that basic language instruction was not what it could or should be.

During a two-year evaluation period many aspects of the language learning experience at Pennsylvania were reviewed. This included analysis of the foreign language learning background of entering students as well as wide scale testing of students enrolled in foreign language courses.

With respect to matriculating students the following were among the data which emerged during this evaluation (Freed, 1984, p. 214):

> 1. Only thirty-three percent (33%) of all College of Arts and Sciences matriculants at the University of Pennsylvania present CEEB Language Achievement Test scores upon entrance. Because students are placed into language classes according to their CEEB scores, this means that some sixty percent (60%) of all incoming freshmen have to be tested during of before the first week of classes. (A small percentage are exempt by virtue of Advanced Placement.)
> 2. Only one-third of those who present CEEB language scores upon entrance receive scores high enough (650 or above) to exempt them from further language study. Even if the minimum score were lowered to 600, no more than half of those students who present CEEB scores would be exempt from further study. In other words, some seventy-five to eighty percent (75-80%) of all entering freshmen have to take course work to satisfy the foreign language requirement.

These data were both surprising and disturbing. We had anticipated that those who felt confident enough to take the CEEB language achievement tests would achieve relatively high scores. We would also have liked to believe that those who had spent valuable years of foreign-language study at the secondary level would have developed skills that would have exempted them from elementary and intermediate-level language courses at the university level. Such was clearly not the case.

The results of an internal testing program which emphasized traditional CEEB Language Achievement Tests were also distressing. The CEEB Language Achievement Tests were administered to all students enrolled in levels 1-4 French, Spanish, Italian, German, Russian, and Latin at the end of the Spring and Fall semesters, 1980. Mean scores for students who were completing the fourth semester of study are shown in

Table 1
MEAN SCORES ON CEEB LANGUAGE ACHIEVEMENT TESTS FOR STUDENTS COMPLETING A LEVEL 4 LANGUAGE COURSE.
(Freed, 1984. p. 215)

French	Spanish	Italian	German	Russian	Latin
558	533	561	575	593	580

Table 1. These data showed that, at the end of the (then) required four semesters of study, our own students were not achieving the 650 minimum score that incoming freshmen needed for exemption from further language study. (For a more complete discussion of these test scores and related data see Freed 1981, 1984.)

These findings reinforced our belief in the need to redefine the nature of the language requirement and further convinced us of the futility of a system which evaluated language acquisition almost exclusively in terms of time and credit units. It had become clear that many students who fulfilled such requirements were unable to use the languages they had studied in any meaningful way. While they had satisfied the demands of the University, they had experienced no personal satisfaction nor had they gained any useful skills. Moreover, there was no convincing evidence that students were particularly well prepared for more advanced courses.

Despite such passionate convictions the transition to a proficiency requirement was not made easily or precipitously. In an environment where student opinion manifestly opposed any foreign language requirement, a change that would appear to make the requirement even more stringent would surely have provoked intense student reaction. In addition, objection to a proficiency requirement was not limited to students. The notion aroused skepticism on both practical and philosophical grounds among certain faculty members. We therefore moved cautiously as we considered several fundamental issues. These included:

1. The definition of an appropriate foreign language requirement as part of an undergraduate liberal arts curriculum.

2. The identification of appropriate minimal levels of functional proficiency for this requirement.

3. The determination of how to test for proficiency.

Perhaps most central to this discussion was the determination of an appropriate foreign language requirement within an undergraduate liberal arts curriculum. Pennsylvania had maintained its foreign language requirement even in the face of student cries for relevance in the 1960's. During this period we capitulated somewhat by changing from a three year to a two year requirement. In the early 1970's a training program for foreign language Teaching Assistants (who teach most of our lower level courses) was introduced (Freed, 1975). This program, which emphasized communicative language teaching, student interaction, negotiation and an input rich environment, was the precursor of the transition to a proficiency-based language requirement. However, while the notion of communicatively-oriented teaching alarmed only the most conservative language faculty, the mere suggestion of changing to a "functional proficiency requirement" was perceived by a small number of other faculty as a concession to the anti-humanistic, anti-intellectual mood of the times. These faculty saw this move as one which threatened the disintegration of our language programs into a series of "Berlitz-like" language courses, courses which stressed only "survival skills," a term with a newly associated pejorative meaning.

The arguments presented by the pro-proficiency contingent were equally forceful. The data collected by the Language Advisory Committee (LAC) clearly demonstrated that the status quo was serving neither student pleas for relevance nor faculty pressure for training students to pursue literary study. Furthermore, those who promoted the proficiency requirement which would be integrated into the self-correcting discipline of departments of language and literature as well as within a university context, and therefore presented little risk of becoming a requirement based on survival skills alone. The LAC developed a statement of purpose of the language requirement which embodied a philosophy toward proficiency-based teaching which in no way sacrificed a general commitment to grammatical accuracy, to the notion of sociolinguistic appropriateness, or to preparation for advanced literary study:

> The language requirement should prepare students to use a foreign language meaningfully. This includes the skills that are required to participate at a minimal level of proficiency in authentic target-culture situations that might call for listening, speaking, reading and/or writing. These skills will, in principle, span a continuum from such day-to-day tasks as giving directions to those skills required in analyzing a work of art. This required level of competence is that which would be minimally acceptable to native speakers and can reasonably be acquired by students in the time normally devoted to required language study. At this level of minimal proficiency, importance is given to grammatical accuracy and the ability to express appropriately one's intent in a variety of situations. . .

This underlying philosophy was eventually adopted by those departments or sections of departments that decided to institute a proficiency-based requirement. However, prior to implementing this redefined language requirement, several additional areas remained to be explored. These included identifying appropriate minimal levels of proficiency to be attained by those satisfying the language requirement and a determination of how to test for proficiency.

Pennsylvania's earliest discussions of redefining the language requirement coincided exactly with the Foreign Service Institute's (FSI) first efforts to introduce proficiency testing and proficiency standards to academic institutions. Representatives of Pennsylvania's faculty attended an early Testing Kit Workshop and others participated in the ETS Common Yardstick Project. It is within this context that our initial explorations took place. Such explorations did not include a theoretical examination of the term "proficiency". Rather, interest focused on how we might go about creating a multi-skill requirement in which students would be required to demonstrate the ability to use language interactively, communicatively and for a variety of purposes. It is important to note here that the original emphasis was placed on standards and not on tests. There was emphatic and explicit understanding that teaching and learning must be directed toward these goals rather than toward the tests. While we knew that students were eager to develop functionally useful oral skills, faculty were reluctant to support a requirement that would exclude broader humanistic goals. Therefore, there was never any question from the beginning that the skills of listening, speaking, reading and writing would be integrated and would be given equal importance in the classroom. It was also understood that all skills would be tested. During the early deliberations the testing of culture was also envisioned, but this was subsequently abandoned as an item to be tested individually.

The standards for speaking proficiency as described by FSI (Adams & Frith, 1979) struck us as a reasonable point of departure. While they had been written for adult foreign service officers and not for undergraduates in American colleges, they provided a framework on which to build the requirement. Consequently, the FSI speaking definitions were initially adopted as descriptions for oral interaction. Similar descriptions were then developed at Pennsylvania for listening, reading and writing.

Identifying the minimal acceptable levels of performance on the proficiency test was essentially a trial and error procedure. There were, however, data to guide us in estimating these levels. These included the findings available from the achievement (CEEB) and oral proficiency testing phases of our own project, the data reported by Carroll in his 1967 study of language majors near graduation from college, and from the language portion of ETS's Global Understanding project (Clark, 1981). These data were used by individual departments to establish standards of performance they deemed acceptable and realistic for students completing the language requirement.

The final major issue at this point was a determination of how to test for proficiency. Once again, because of general dissatisfaction with the requirement then in existence and the underlying philosophy that it reflected, interest focused on means of testing oral interaction and listening comprehension (skills previously neglected) as well as testing reading and writing. Our recent experience in the FSI workshops created considerable albeit not unanimous enthusiasm for the oral proficiency interview. Despite our realization that this would require an unprecedented investment of time in training and testing, the oral proficiency interview was adopted as the means of testing oral interaction. Each department then undertook the task of developing new proficiency-oriented tests for listening, reading and writing which were to be pilot tested in the coming years.

Deliberation of these issues culminated in September 1981 when the French section of the Department of Romance Languages decided to replace the traditional language requirement by one which would be satisfied by a proficiency examination to be taken after several courses of study.[1] In September 1982, German, Russian and Arabic instituted a similar requirement. Italian followed in 1983 and in 1984 the Department of South Asia Regional Studies began to explore the feasibility of a proficiency-based requirement. It is expected that Spanish will join other Romance languages in the fall of 1987.

While it is not within the scope of this paper to provide a detailed description of the proficiency requirement in each department, some background information regarding expected levels of proficiency, means of testing, and general policies in administering the requirement is pertinent.

The proficiency examination is given at the end of each semester to those students who are completing the fourth semester (or the equivalent) of study.[2] However, students who wish to take the examination earlier (e.g., during the third semester) are permitted to do so. Students must pass the examination in order to satisfy the language requirement. Therefore students who pass the examination after three semesters are able to complete the requirement early. By contrast those who fail the exam after four semesters have not satisfied the language requirement despite the fact that they have been enrolled in language courses for the equivalent of four semesters. Students who are unable to pass the proficiency test are counseled on an individual basis. They are then given the option of repeating a course, or studying either on their own or with the assistance of a tutor. In principle they are able to retake the exam until they pass it. However, exceptions are sometimes made in the rare circumstance where a student is unable to pass the proficiency test after several efforts.

The policies for implementing the proficiency requirement are as follows:

> 1. The proficiency test consists of five separate sections. Proficiency is determined by computing a composite score for all portions of the test. A student is thus able to compensate for weakness in one area by above average performance in another.
> 2. Students must receive a composite passing score on the proficiency test in order to be eligible for a grade in the level four course. No letter grade is assigned on the test itself. Results are recorded as either pass or fail.
> 3. Students who have passed the test are given a course grade based on their cumulative performance throughout the semester. This grade is based on quizzes, homework, hour exams, etc. The form and content of many of these tests and assignments is similar to areas tested in the proficiency test.
> 4. Any relationship between the scoring system and A,B,C,D grades has been eliminated by using numerical categories for rating the proficiency test.
> 5. Students who fail the proficiency test are required to improve their performance on the portions of the test they failed, or raise their score on passed portions to achieve a cumulative passing average.

Table 2

MINIMAL PASSING PROFICIENCY LEVELS

Language:	Arabic	French	German	Russian
Skill				
Oral Interaction	Int-Mid	Int-Mid	Int-Mid	Int-Low
Listening Comprehension	Int-Mid	Int-High	Int-High	Int-Mid
Reading Comprehension	Int-High	Advanced	Advanced	Int-Mid
Writing	Int-Low	Int-High	Int-High	Int-Low
CEEB	NA	500	500	NA

The ACTFL Guidelines have now replaced the FSI descriptions and have been informally adopted by individual departments as a general organizing format to provide common standards across departments and to define a range of identifiable levels. It is important to stress here that instruction is neither bound nor limited to items in the Guidelines. Each language has identified various levels as those which represent the minimal level necessary to demonstrate passing performance. Instruction clearly aims at higher levels and students are unaware of the minimal required level. The table below illustrates the minimally acceptable level in four language sections for each of the skills mentioned above. For illustrative purposes, standards are presented for Arabic, French, German, and Russian.

This table portrays the most current description of minimally acceptable standards for each section of the proficiency test. This description represents considerable revision from the levels initially specified in several languages. For example, both German and French originally required Intermediate-High as the minimal passing level for oral interaction; both however found that this level was unrealistically high and have substituted Intermediate-Mid in speaking as the minimal passing score. Obviously many students do score at (and sometimes above) Intermediate-High, but some students, particularly those who come from grammar-oriented high school classes, are unable to achieve the Intermediate-High level within one or two college semesters.

While the details of the proficiency requirement vary from department to department, the principle remains the same. Teaching is oriented towards helping students acquire functionally useful communicative abilities as well as structural accuracy. As previously stated, the objective of our language courses and of the proficiency test is not to equip students with mere survival skills but rather to provide students with the ability to use language appropriately for some meaningful purpose in a variety of situational contexts. These purposes will vary from student to student and from time to time. The goal of the requirement is to prepare students for more advanced language study as well as to provide some usable skills for those who will discontinue language study once they have satisfied the requirement.

The proficiency test examines skills in oral interaction, listening comprehension, reading comprehension and writing. In addition, the College Entrance Examination Board Language Achievement Test (CEEB) is given to students in most language sections as a fifth part of the proficiency test.

The means of testing each skill also vary slightly from language to language. In all cases, the ACTFL adapted Oral Proficiency Interview (OPI) is used to measure oral interaction. Listening comprehension tends to be measured by asking students to provide a resumé of a casual conversation between two native speakers or some other authentic natural recording. Evaluation of listening comprehension is based on both accuracy and the extent of comprehension. Reading comprehension tends to be measured by the student's ability to provide a resumé of the most salient features of an authentic non-literary and sometimes also a literary text. As in the case of listening comprehension, evaluation is based on both accuracy and the quantity of detail. Tests of writing ability are similar to guided compositions. Students are asked to address a general topic, frequently in the form of a letter. Evaluation is based on the ability to carry out the task, appropriate language use, vocabulary and functional proficiency as well as grammatical accuracy.[3]

The foreign language proficiency requirement has now been in place for several years. While philosophical issues were settled early on, numerous other issues, both practical and administrative, have remained which have had to be dealt with on a continous basis. These include the training and supervision of those who teach proficiency-based courses, the choice of materials for these courses, counseling of students, and administrative decisions regarding interpretation of the requirement.

The training and supervision of Teaching Assistants has been a regular component of our foreign language programs for many years. However, the implementation of the proficiency requirement imposed new demands for T.A. time and training. In addition to a pre-semester orientation program which emphasizes a communicative approach to language teaching and required pedagogical seminars (for graduate credit in some though not all departments), special workshops in teaching and testing for oral proficiency have had to be organized. These include annual introductory workshops on oral proficiency testing followed by intensive three-day testing workshops held at mid-semester. These workshops are conducted by Pennsylvania faculty who are certified ACTFL testers and trainers. Teaching Assistants are encouraged though not required to complete the ACTFL certification process. In fact, relatively few ever find the time to do so, a problematic concern nationally even more than at Pennsylvania. The increased time demands expected of all language instructors in requirement level courses fortunately has not become a problem. It is a fact accepted by all those who accept positions, that the time required for both training and testing is an expected part of employment.

Practical considerations in establishing the proficiency-based requirement have extended to materials development as well. In instituting a proficiency requirement it was implicitly understood that classroom instruction would have to reflect the standards established. A communicative and interactive approach to language teaching had long been encouraged but this redefinition of the requirement brought an added urgency to reevaluation of course texts and materials. New textbooks and materials were adopted by language sections in several departments. Under the auspices of the Regional Center for Language Proficiency efforts were made to define proficiency-based activities (Freed, 1984) and to develop classroom materials which would integrate various language skills, focus on natural discourse and help students develop functionally useful skills. The creation of authentic materials for use in the proficiency-oriented classroom remains an ongoing concern.

Of equal importance is the counseling of students in language requirement level courses. In the early years of the requirement there was a great deal of student apprehension. Students perceived the redefinition of the standing requirement as an overwhelming threat and many tried (in vain) to be exempted from what they considered a retroactive requirement. As the proficiency requirement became an established part of the undergraduate curriculums these responses have essentially disappeared. Nonetheless, the proficiency requirement has created a need for more careful student guidance and evaluation. Students who wait until their senior year to complete the language requirement place themselves in great jeopardy. Failing the proficiency examination during a student's final semester can prevent graduation. Students who fail the examination, at any point in their undergraduate career, must be advised as to the options available to them for improving their performance before retaking the exam.

The institution of the proficiency requirement brought with it certain administrative problems and obligations. The most obvious and perplexing have been largely resolved. These related to the interpretation of a requirement which appeared to many students as retroactive. The decision to consider successful performance on the proficiency test as a prerequisite to completing the fourth semester of language study made this a departmental initiative, rather than a college requirement. It was therefore decided that students would not be permited to take advantage of a Grandfather's Clause. It would however be misleading not to admit that greater leniency existed in evaluating the proficiency examinations during the first two years than is the case today.

There has also been an administrative obligation to be certain that timely descriptions of the various means of satisfying the language requirement appear in print. This is particularly so in the situation where many but not all languages have elected to establish a proficiency requirement.

This paper has been written with multiple intents: to provide a description of the issues and problems encountered by the University of Pennsylvania in establishing and maintaining a foreign language proficiency requirement, to suggest to others who would consider a similar move what issues and problems must be anticipated and, finally, to provide a specific context for evaluating the application of the ACTFL Proficiency Guidelines and the ILR-ACTFL oral proficiency interview in an academic setting. To review, two general categories of concern have been identified: philosophical and practical. The philosophic issues relate primarily to the definition of an appropriate foreign language requirement for an undergraduate liberal arts curriculum at the end of the twentieth century. On a more concrete level they include decisions regarding the identification of appropriate minimal levels of functional proficiency for such a requirement and decisions regarding testing mechanisms. From a purely practical perspective several other issues are apparent. These include the training and supervision of those responsible for proficiency-based courses, the selection and development of materials for these courses, the counseling of students who study at this level, and administrative decisions regarding the interpretation of the requirement.

There is a final area of concern not previously discussed which does not fit neatly into the other two categories. This is the fact that specific aspects of the proficiency requirement must remain in a constant state of revision, sensitive to new needs and emerging data. This does not imply that the notion of a proficiency requirement is here subject to question, but rather that the definition of proficiency itself, of required levels of performance, of materials and tests, must continuously be revised and refined.

Insightful and penetrating criticism of the ILR-ACTFL Guidelines and tests has recently emerged (Bachman & Savignon, 1986; Bachman, 1987; Jarvis, 1986; Kramsch, 1986 a & b, 1987; Schulz, 1986; Savignon, 1985, 1987) and more will surely be forthcoming. It would be not only premature but emphatically foolish for any organization or institution to slavishly adhere to the earliest products of the proficiency movement. It would be equally foolish to disregard or destroy the valuable impact this body of work has had on many aspects of language learning and language teaching.

From the perspective of many language faculty at the University of Pennsylvania, the commitment to proficiency-based instruction has improved the language learning environment at Pennsylvania. It has fostered creativity, promoted a more integrative approach to language instruction and helped improve articulation between lower level courses. Moreover, the thoughtful use of the ACTFL Guidelines within the Pennsylvania context has had a liberating effect in terms of expanding and enriching the foreign language curriculum. Rather than limiting the available options for instruction, the adoption of a proficiency-based requirement has freed us from the tyranny of grammatically organized textbooks and a legacy of traditional approaches to language teaching.

There is no question that the investment in instructor energy in terms of training, teaching and testing has been enormous, but the results seem to make it worthwhile. It would, however, be misleading to imply that we have data to demonstrate that proficiency-based instruction has actually improved language learning. The fact is that with the exception of CEEB scores which have improved slightly over the years, there is no available pre- and post- proficiency test data. The fact is that the length of required time has not changed radically for most students. Some students do in fact complete the language requirement early and others require a fifth semester. However, the majority of students still spend the equivalent of four semesters in required study. Only limited proficiency can be achieved in this period of time, regardless of how proficiency is defined. We are anxious to learn what the long-term effects of proficiency-oriented courses will be with respect to continued motivation, preparation for more advanced courses, and retention of material learned. Impressionistically we are aware of few complaints about the language requirement now that the proficiency requirement is a well established and integral part of the curriculum. While it is unlikely that the foreign language requirement will ever be the most popular aspect of the undergraduate requirement, the observations of students and faculty alike regarding increased motivation and greater inherent interest in course content make it, in our opinion, more than worth the effort.

In conclusion, we at Pennsylvania are encouraged by what we have seen thus far. However, it is more obvious than ever that the emphasis must be on the creation of realistic goals for language learning as well as an understanding of the implications of these goals and standards by all those involved in the language learning process. To the extent that the emphasis is on tests and testing, real changes will not necessarily occur. In fact, it could even happen that the term proficiency standards or proficiency requirements could become as empty as were the former catalogue descriptions that invoked the term "competence" but really meant "time". Perhaps more than ever before we must provide for and maintain an intimate connection between training, teaching and testing.

Notes

[1] The information contained in the first and last sections of this paper is a revised version of material to appear in Freed (1987).

[2] It should be understood that all students do not necessarily study language at Pennsylvania during four consecutive semesters. Many students are placed into language levels 2, 3, or 4 upon entrance. Initial placement into language courses is based on scores on a retired version of the College Entrance Examination Board Language Achievement Test (CEEB) which is used as a placement instrument.

[3] It is important to emphasize that the testing mechanisms, as well as minimal passing levels, are constantly being revised and refined. While the notion of a proficiency-based requirement is totally endorsed, the definition of proficiency, and implications for teaching and testing toward the communicative use of language, will continue to change.

References

Adams, M., & Frith, J. (Eds.). (1979). *Testing Kit: French and Spanish.* Washington, DC: Foreign Service Institute.

Buchman, L., & Savignon, S. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. *Modern Language Journal, 70*, 380-390.

Bachman, L. (1987). Problems in examining the validity of the ACTFL oral proficiency interview. This volume.

Carroll, J. B. (1967). Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals, 1*, 31-51.

Clark, J. D. (1981). Survey measures: Languages. In T. S. Barrows (Ed.), *College students' knowledge and beliefs: A survey of global understanding* (p. 32). New Rochelle, NY: Change Magazine Press.

Freed, B. (1975). Why train teaching assistants? *ADFL Bulletin*. 7, 9-14.

Freed, B. (1981). Establishing proficiency-based language requirements. *ADFL Bulletin*, 13, 6-12.

Freed, B. (1984a). Proficiency in context: The pennsylvania experience. In S. Savignon & M. Berns (Eds.), *Initiatives in communicative language teaching* (pp. 211-240). Reading, MA: Addison-Wesley.

Freed, B. (1984b). What makes an activity proficiency-based? Paper presented at the ACTFL Conference, Chicago, IL.

Freed, B. (1987). Preliminary impressions of the effects of a proficiency-based language requirement. *Foreign Language Annals*, 20, 139-146.

Jarvis, G. (1986). Proficiency testing: A matter of false hopes? *ADFL Bulletin*, 18, 20-21.

Kramsch, C. (1986a). From language proficiency to interactional competence. *Modern Language Journal, 70*, 366-372.

Kramsch, C. (1986b). Proficiency versus achievement: Reflections on the proficiency movement. *ADFL Journal, 18*, 22-24.

Kramsch, C. (1987). Response to H. Byrnes. Discourse structure and communicative strategies in the development of communicative ability. This volume.

Savignon, S. (1985). Evaluation of communicative competence: The ACTFL provisional proficiency guidelines. *Modern Language Journal, 69*, 129-134.

Savignon, S. (1987). Response to B. Freed & D. Lange. Practical problems in the administration of a proficiency requirement and in the implementation of the ILR-ACTFL proficiency test and guidelines at the high school and university level. This volume.

Schulz, R. A. (1986). From achievement to proficiency through classroom instruction: Some caveats. *Modern Language Journal, 70*, 373-379.

DEVELOPING AND IMPLEMENTING PROFICIENCY ORIENTED TESTS FOR A NEW LANGUAGE REQUIREMENT AT THE UNIVERSITY OF MINNESOTA: ISSUES AND PROBLEMS FOR IMPLEMENTING THE ACTFL/ETS/ILR PROFICIENCY GUIDELINES

Dale L. Lange
University of Minnesota-Twin Cities Campus

This article discusses a proficiency-based foreign language requirement at the University of Minnesota and the testing program which has been developed to respond to it. The requirement indicates that different levels of competence in language are desired upon entrance and mandated for graduation. Tests for the evaluation of competence in listening, reading, speaking, and writing have been developed over a two-year period. They have been pilot tested in University of Minnesota courses and in public schools and administered for the first time in Fall, 1986. The article examines the process of test development and pilot testing, discusses some of the perceived strengths and limitations in the implementation of the ACTFL/ETS/ILR proficiency guidelines, as well as some general concerns about the nature of language learning in higher education.

1. Introduction

Interest in the teaching and learning of foreign languages is currently very high. Colleges and universities are reinstating language requirements in efforts to "improve the quality of undergraduate education." State departments of education are re-examining their requirements for elementary and secondary education. In some cases, this situation has brought new statements about the amount of time foreign language should be given in the curriculum. Other arguments attempt to demonstrate the importance of languages to the "national interest: the economy and foreign trade, politics and international diplomacy, and security. The results of this renewed interest are demonstrated as follows: Enrollments in languages are up in both schools and universities. In some parts of the country, there are severe shortages of teachers; new programs are being discussed. There is vigor again in the discussions about language learning and teaching. And, the renewal of foreign languages has brought with it some different issues for our attention, new problems for resolution, and interesting challenges for our consideration. In this context, the focus of attention is on what can be done with language. Will students at the end of a program or as the result of a requirement own some demonstrated competence? Will they be able to communicate beyond simple requests for food, clothing, and shelter? Can they express an idea in formal written language? Can they use the language for academic purposes?

The intent of this article is to relate to some of the very practical problems that have arisen in the implementation of the ACTFL/ETS/ILR Guidelines in a new language requirement at the University of Minnesota. This major university is serious about the need for language study in its continuing international role. Particularly within the liberal arts tradition, the university wishes to prepare students to use language as a means for exploring the world of ideas, thought, and the research of others in other languages to better understand their own. As a result, this new requirement is oriented to a functional use of the language and away from seat time.

2. A new language requirement based on competence

Since 1970 and until Fall Quarter of 1986, students at the University of Minnesota completed a language requirement in one of two ways:

1) either they took five quarters of language; or 2) they took three quarters of language and three quarters of culture courses related to the language they have been studying, the "scenic route." For example, student X may have taken five quarters of German to complete the requirement. Student Y may have taken three quarters of French and then a course in 17th century French painting, 18th century French history, and 20th century French music for the same purpose. Significant differences in outcomes seemed to have been achieved from the two approaches to the former requirement. Most students were opting out of the five quarter sequence of language courses for the three quarters of language and three quarters of culture courses. And faculty members in many departments were concerned that students could not use their competence in a second language for research purposes.

In January 1983, the Dean of the College of Liberal Arts appointed a Taskforce on Language Instruction to examine the then existing language requirement. The responsibility of this Taskforce was to examine the situation and make recommendations for necessary changes to correct the problems. By June of 1983, the Taskforce made the following proposals:

1. The two different routes to the fulfillment of the language requirement would end.

2. Students who would enter the College of Liberal Arts and who intended to complete a Bachelor of Arts or Bachelor of Fine Arts degree would meet an Entrance Standard in a second language in listening, reading, speaking, and writing and a Graduation Requirement in the same modalities.

3. In either the Entrance Standard or the Graduation Requirement, students would demonstrate competence in the language rather than an amount of time spent in class as a means of meeting the requirement. Officially, the requirement states that students present the "proficiency usually attained after three quarters of college study" for the Entrance Standard. For the Graduation Requirement, students would need to demonstrate the "proficiency...usually attained after six quarters of study."[1]

4. The requirement becomes fully effective, Fall Quarter, 1988, with a phase-in period beginning Fall Quarter, 1986, for the Entrance Standard. In the phase-in period, admitted students must present the equivalent of one quarter of language proficiency study in Fall Quarter, 1986, the equivalent of two quarters in 1987, with the full requirement

coming into existence in Fall Quarter, 1988. Another three quarters of language study would be offered to help students meet the Graduation Requirement. Tests to be developed for use in the Entrance Standard and Graduation Requirement would examine designated competencies in all language modalities.

5. French, German, and Spanish, because they are the languages most taught in secondary schools in the state of Minnesota, are the primary languages. If students do not possess the appropriate competence to meet the Entrance Standard in these languages, they must enroll in an appropriate course to help them achieve the needed competence. However, they receive no credit for these courses for graduation. As an alternative, students may begin the study of less-commonly taught languages, such as Arabic, Chinese, Japanese, or Russian, for credit.

6. The less-commonly-taught languages, must develop proficiency tests by 1990 to test proficiency levels that they will determine.

In March of 1984, the College of Liberal Arts Faculty Assembly voted, without dissent, for these changes. The next issue would be the implementation of the policy changes in the language requirement, certainly no small task since there were no extant proficiency oriented tests and since most curricula in secondary schools and at the University were oriented toward the grammatical syllabus. For a more detailed description of the new language requirement and its initial implementation, see Arendt, Lange, and Wakefield (1986).

3. Implementation of the new language requirement

The first step in implementing the new requirement involved reaching an understanding of the importance of the new policy and the necessity for cooperation among different levels of the education system in the state of Minnesota. For this purpose, an oral proficiency workshop was held in the summer of 1984. This workshop brought together 30 participants whose tuition for the course was paid by the University and other outside funds. Three languages were represented: French, German, and Spanish. Fifteen of the participants were secondary school teachers and fifteen were college/university faculty, some from the University of Minnesota and some from private colleges, the state university system, and from community colleges in the immediate area of Minneapolis/St. Paul. This workshop initiated the group to the ACTFL Provisional Proficiency Guidelines (1982), the oral interview procedure, and implications of the guidelines and interview procedure for instruction, curriculum development, and the testing of other modalities (listening, reading, and writing). The result of this experience was the development of a cadre of people who were enthusiastic about the prospect of change in the focus of language programs.

The second step included three implementation workshops in the winter and spring of 1985: 1) definition of curriculum; 2) definition of test parameters and constraints; and 3) item writing. The University enlisted the cadre of people prepared during the Summer of 1984 for this purpose. There was a balance of secondary school and college/university people for these experiences. Each of these workshops is defined briefly.

Definition of curriculum – Two difficult tasks faced this group: the designation of expected levels of proficiency for both the Entrance Standard and the Graduation Requirement. After much discussion and negotiation, the first task was completed. The levels were set for these three languages as follows:

	Entrance Standard	Graduation Requirement
Listening:	Intermediate-Low	Intermediate-High
Reading:	Intermediate-Low	Intermediate-High
Speaking:	Novice-High	Intermediate-Mid
Writing:	Novice-High	Intermediate-Mid

With these designated proficiencies, participants proceeded to the second task. They wrote specific learner outcomes each for French, German, and Spanish. These outcomes are based on the concepts of full and partial control of content, function, and accuracy as related by Heilenman and Kaplan (1985). The negative nature of the guidelines was also changed. Statements about what students could not do were changed to what they could be expected to sustain at the level designated and what they might be able to perform at least partially beyond that level. Each language modality was given the same treatment. The document produced as a result of this workshop for each language serves as a framework for curriculum development within the University language departments (University of Minnesota, 1986). It is on this same framework that the testing program is constructed. Further, it serves as a communication with secondary schools, community colleges, and other colleges/universities in the state.

Definition of test parameters and constraints – This second workshop examined the realities and the constraints of testing 1,500-2,000 persons for the Entrance Standard alone. It also examined different kinds of test items, their usefulness, efficiency, and effectiveness in evaluating the desired competence. The form and shape of the tests were also determined.

Item writing – In the final exercise, twenty-five participants developed a pool of items for both the Entrance Standard and the Graduation Requirement for the three languages and the four modalities. In the second week in June, 1985, these participants wrote hundreds of items to meet these needs.

Described by Lange and Lowe (1986), the guidelines for reading, in consort with a set of 11 texts which had been rated by ILR staff, were used in an initial exercise for item writing. Participants ranked the passages from easiest to most difficult and then rated them according to the ACTFL Guidelines without the benefit of knowing how ILR staff had either ranked or rated the same passages. The exercise allowed the participants an opportunity to internalize the levels of the guidelines. After discussions of the nature, qualities, and approaches to writing objective type items in the testing of listening and reading and after examining the kinds of tasks that needed to be developed for speaking and writing, participants wrote items for a test item bank for each modality in the remaining four days. These items served as the basis for the listening, reading, speaking, writing tests that were developed in the 1985-86 academic year.

The third step in the implementation process was the development of the tests for both the Entrance Standard and the Graduation Requirement in the four language modalities. The tests were constructed throughout the 1985-86 academic year. They follow the form of an oral interview with the intent of determining the the students ability to *sustain* the language behavior described in the levels chosen for each modality. Each of the tests is organized in four parts: 1) warm-up (items below the determined level); 2) level-check (items at the determined level); 3) probe (items above the determined level); and 4) wind-down (items at the level-check or warm-up level). Level-check and probe items are generally mixed. The test is thus structured to measure the students ability to sustain the level determined since 80% of the items are at level-check or below (warm-up 20%, level-check 40%, wind-down 20%). The twenty percent of items at the probe level indicate the student's ability to function beyond the determined level. A description of the tests is provided in the Appendix.

In Fall Quarter, 1985, test development began. The items from the early summer were all read; many items were eliminated; others were severely edited. Because it was not possible to develop all tests simultaneously, the reading tests were constructed first. Both Entrance Standard and Graduation Requirement tests were piloted with several sections of students in French, German, and Spanish at the end of the quarter. The Entrance Standard tests in the three languages were also piloted with secondary school students in one of the larger suburban school districts of Minneapolis during the winter of 1986. The reading tests were given again in the winter and spring quarters, 1986, but with university students only.

The listening tests for Entrance Standard and Graduation Requirement in all three languages were developed and piloted during Winter Quarter, 1986. Only university students were used in this exercise. These tests were also given during the spring quarter. The writing and speaking tests for both Entrance Standard and Graduation Requirement were piloted in Spring Quarter, 1986. Only one speaking test was written, namely that for the Entrance Standard. It is similar to the Defense Language Institute experimental ROPE (Recorded Oral Proficiency Examination). A tentative decision to give oral interviews for the graduation requirement eliminated the necessity for a second speaking test.

The pilot test results in listening and reading gave us an indication of the difficulty and discriminability of items for those tests. These data were used to examine items and revise or eliminate them. The reliability of the tests, generally above .70, was sufficient to indicate usefulness of the tests. Results with students showed differences by quarter for university students in listening and reading and by year for secondary secondary students in reading. Pilot tests results in writing showed us that, with a four point rating system each for content, function, and accuracy, interrater-reliability could be achieved (.95+). From a limited sample of students, speaking test pilot results indicated that a recorded oral proficiency evaluation would work.

The tests and the new language requirement were implemented for the first time at the beginning of the Fall Quarter, 1986.

4. Issues and problems in the implementation of the requirement

The development of tests for large scale testing purposes such as a new language requirement, based on the demonstration of competence, demands a recognition of test constraints and other problems which face those initiating such a program. These issues are discussed below.

There were at least eight categories of constraints that we struggled with in preparing tests for 2,000 students. They are as follows: efficiency, number of test forms, authenticity of language, single or mixed modalities, item validity, scoring,

Efficiency – The tests to examine language competence had to be developed with as little cost as possible. That principle indicates that students would have to be tested in large numbers when possible (Example: writing and reading); scoring would have to be by machine, when feasible (Example: listening and reading); when scoring by machine was not possible, language department TAs would be required to score tests (Example: speaking and writing). Testing would be limited to one location, namely the University of Minnesota campus. Because large numbers of students would be involved in the testing program, oral interviews would be impossible to give, especially for the Entrance Standard. For this reason, the ROPE type oral test was developed. A major constraint, problem, and concern in this category is the preparation of teaching assistants to understand and use the guidelines. While their role is mainly in the scoring of the speaking and writing tests, they must still understand the guidelines for those modalities and apply that understanding to the scoring of the tests.

In one way or another, all of these matters relate to the expenditure of both time and money. Efficiency in the use of time and money remain a continuing concern in our test development endeavors.

Number of test forms – In order to make the tests secure, three forms of each test would be required by 1988 when the language requirement becomes fully effective. But because of severe strain on budget, time, and staff, only one form of each of the test exists currently. But we intend to build a computerized bank of items for each test. That project begins during spring quarter, 1987. The item-bank computer program allows items to be drawn from it to create a test in almost any shape required; it also allows items to be retired from a test created with it. Those items will be stored and used in later versions of the tests. The computer program will help us prepare tests for listening, reading, and writing in the shape of the oral interview when it is told to do so. This test item bank will alleviate concern for test security, but it is a costly process because of the need for personnel to input items into the system.

Authenticity of language – This constraint effects every test in one way or another, but especially those for speaking, reading, and listening. The writing test stimuli are in English; they are not of concern here. However, for speaking, listening, and reading, authenticity of language is a serious problem particularly at the lower levels. For speaking, it is mainly the supply and use of authentic speakers in the target language. Restricted topics at Novice-Mid, Novice-High, and Intermediate-Low on a recorded interview is also an issue. The recorded nature of the test is yet another concern. For reading, there are few authentic texts at the Intermediate-Mid and Intermediate-High levels. Either they need editing or the questions asked of students to demonstrate their comprehension must be targeted to the determined level. Listening texts are much more of a problem. At the designated levels of competence for the Entrance Standard in particular, we have had to construct them, largely with the help of native speakers. Such texts are unnatural, even though written by native speakers, because they were especially constructed for test purposes. Higher level authentic listening texts are less problematic, but also difficult to find. They can come from short-wave, films, videotape soundtracks, etc. But whether for listening or reading, no text is of a single "proficiency" level. Students also bring their own experience to text comprehension with either listening or reading, further confounding what appears to be a clean relationship between choice of text and proficiency level, but which clearly is not.

Single or mixed modalities – In reality, language modalities are not clearly and distinctly separated. However for this requirement, it was decided to to test each modality separately to determine "specific" competence. It is from this decision that the four tests for the Entrance Standard and the three tests plus the oral interview for the Graduation Requirement were constructed. In order that these tests be "pure," that is to allow students opportunities to show what they could produce or comprehend in the language, two other decisions were made. The directions would be offered in English so as not to confuse students responses with lack of comprehension of the test directions. And for the listening and reading tests, all questions and distractors would be provided in English so that the comprehension of the texts would not be compromised by questions and distractors in the target language. This constraint is more related to those outside the project who see this as a major compromise to some approaches to teaching. Yet, students come to the University of Minnesota from many different teachers and programs. We want them all to have a fair chance on the tests.

Item and test validity – The constraints in this category are related to the nature of testing itself. To test competence in a language modality, a test and its items should resemble normal tasks for that modality in so far as possible. While that principle may be more applicable to the production modalities (speaking and writing), with items that ask for responses to direct questions, statements, and situations, it is not so applicable to the reception modalities (reading and writing). Here, items to determine competence are *not* direct, but indirect. Traditionally, the form of such items are multiple-choice, true-false, matching, and fill-in such as in cloze tests. These kinds of items do not relate to the normal tasks of listening and reading, but rather to behaviors from which the testers judge competence. Multiple choice items were chosen finally because they are the most efficient means of determining indirect competence. They could be scored by machine. Measures such as the immediate recall protocall were judged inefficient because of the difficulty in scoring procedures.

In terms of a hierarchy of tests conforming to the principle stated above, the writing test is the most "face valid!" Although recorded, the speaking test, providing a somewhat unnatural environment for this task,, does ask students to respond using the language in normal ways. The reading and listening tests are less face valid because they are indirect measures of competence in those modalities. All of the tests have been constructed from the descriptions in the guidelines.

Scoring – The tests for each modality are to be scored in the most efficient and appropriate manner for the item type. Objective means have been used for listening and reading. A rating system is used for speaking and writing. It evaluates the students competence on each item for content, function, and accuracy: 0=no functioning; 1=below designated level; 2=at designated level; 3=above designated level. This rating system is more sensitive to student responses and is somewhat inefficient, while objective scoring is more efficient in use of both resources and time.

Administration – Like all other aspects of the testing program, test administration must be responsive to available resources. Several approaches to test administration were examined for incoming freshmen:

1) the tests would be given to students in their own schools at the end of the academic year; 2) centers for testing would be established in different parts of the state; students would be tested in these centers at the end of the academic year; 3) students would be tested from August through September as they appeared on campus for a two day orientation to the University of Minnesota; 4) students would be tested, en masse, during the "Welcome" week just prior to the beginning of fall quarter. It was decided that students entering in a fall quarter would be tested during the "Welcome" week just prior to the beginning of fall quarter, 1986 for both logistical and financial reasons. And, students entering in other quarters of the academic year would be tested prior to admission, during the one day orientation days between quarters. These decisions made sense for 1986-87 since the new language laboratory was only installed in October of 1986, allowing only reading and writing to be tested. A new schedule will be constructed to include all modalities for both the Entrance Standard and the Graduation Requirement. For test administration, place and availability of the largest number of students at one time are key factors of efficiency.

Power tests or speeded tests – The Entrance Standard and Graduation Requirement tests for listening reading, speaking, and writing were not developed specifically as timed or speeded tests. We have attempted to give students reasonable amounts of time for the purpose of completing the tests and scoring well. However, we must also limit time to insure efficiency of test administration. In our pilot test experience, approximately 90% of students finished the reading tests in 35 minutes and the writing tests in 30 minutes. The listening tests are timed and vary between 25-30 minutes depending on level and language. The Entrance Standard recorded speaking test is about 10 minutes long, depending again on language. The Graduation Requirement speaking test is an oral interview of approximately 15 minutes. The total test battery is slightly less than two hours for either the Entrance Standard or and Graduation Requirement. However, concern for the length of the test batter is forcing us to consider the elimination of test items and the amount of time used.

These are some of the major constraints that face this testing program. In each one of them, the factors of time, money, and personnel are always present.

5. Other problems relating to implementation of the guidelines

There are several more general categories of problems other than testing which face the implementation of the guidelines at the university level. Some of them are highlighted here. They are as follows: the nature of the ACTFL/ETS/ILR guidelines; the nature of language learning on the university level; and articulation.

The ACTFL/ETS/ILR Proficiency Guidelines – As a description of language behaviors, the guidelines are not complete. Based on experience, rather than a research base, the guidelines are generally flawed in favor of what appears to be language behavior; they are not connected to models of development for each modality. There is an experiential base, but no theoretical or research bases to support them. For example, there is no explicitly stated model of reading in a second language, such as that of Bernhardt (1986), upon which the guidelines are built. This flaw is important and needs resolution in the long run. Bernhardt's (1986) very excellent paper on reading in a foreign language points to the reconciliation of second language with first language reading. The demonstration of a constructivist model based on research with the immediate recall protocol is an extremely important contribution to understanding the nature of reading in a second language. This model of reading in a second language is not yet very well synchronized with the current means of evaluating student comprehension of second language texts. It appears that there is no easy solution to the evaluation of competence in reading comprehension with any of the means suggested, either the "truly objective" such as MC/TF, CLOZE, or the more open-ended recall protocol. The theory has simply advanced beyond practice at this point. The resulting tension allows discussion of both theory and practice. It requires us to focus our testing research agenda toward finding a resolution.

One of the most serious problems for the implementation of the guidelines is the avoidance of statements relating to the socio-cultural context until the Intermediate-High/ Advanced levels (ACTFL, 1987). The elimination of the culture guidelines in the publication of the recent generic guidelines and the provision of no other conceptualization suggests, to those who wish to make the interpretation, that the socio-cultural context contributes little to language development in these earlier stages. Crawford-Lange and Lange (1984), in examining and explaining the use of a process which combines language and the socio-cultural context in language learning, argue that the socio-cultural content is as important as language in the development of linguistic proficiency. If the guidelines do provide a broad framework for the development of language competence (Galloway, 1987), then further evidence of the importance of the socio-cultural context must appear in some form. Without further conceptualization and agreement on this issue, curriculum and materials development, classroom instruction, and testing will tend to avoid the inclusion of the socio-cultural context.

A particular problem for the use of the guidelines in secondary and elementary schools is that they were organized from the observation of adult competency. They do not relate to the development of verbal abilities, functioning, and content of children and adolescents. Without some recognition of the differences, as well as the similarities in language development between children, adolescents, and adults, the guidelines could be easily misinterpreted as being applicable to students on any level, especially by those who are unthinking "bandwagoneers." The guidelines are biased in favor of the "well educated native speaker." The use of the guidelines with minority populations in the U.S., in particular speakers of Spanish, suggests to them once again that their language, particularly its spoken form, is inferior. While some of the bias is removed in the latest ACTFL version with less emphasis on the WENS, the problem remains. Those implementing either a curriculum or a testing program must account for this issue. The current nature and form of the guidelines make their implementation problematic only if they were to be changed too dramatically and too often. Further, such implementation will not occur if decision-makers believe that they are not "theoretically" pure. At the University of Minnesota, for example, we recognize the arguments both pro and con concerning the nature of the guidelines as well as their several flaws. However, we make no apologies for using the ACTFL Guidelines; we decided to use them in spite of the flaws. The ACTFL Guidelines are the only complete system upon which we could build our requirement. The decision to use them was a conscious one; we did it with open eyes.

The nature of language learning in universities – While it is dangerous to be categorical in describing language learning in universities, it is also important to be honest. There are conditions which impede the implementation of the guidelines on the university level. Many of these issues are important to language programs and language instruction *without* considering the implementation of proficiency guidelines. Some of these matters are discussed below:

1. Teaching staff – The teaching staff at the university level is composed largely of teaching assistants who are supported to pursue literary or linguistic studies at the doctoral level. As a matter of course, they are not necessarily disposed to language teaching other than as a job. Language coordinators attempt to prepare the TA instructors in effective instruction, curriculum development, materials development, and testing in a short period of time prior to the onset of the academic year. Further time is generally not available or taken. In this model, is it tacitly assumed that the TA as native speaker or the TA with advanced competence, because they "know" the language, are already quality instructors even without teaching experience. Because the basic assumption here is false, and because these students will teach as much language as their literary or linguistic specialty after their studies are finished, further attention must be directed toward teacher preparation and evaluation in the implementation of the guidelines if successful use of this framework is intended.

2. Program coordination – The quality of program coordination in large university programs acts directly on student outcomes. The coordinator affects the quality of the program in the preparation and supervision given to teaching assistants, the direction given to the curriculum, and the plan for student evaluation. The preparation of these individuals and the reward system to continue their support once hired is crucial to a carefully articulated program. Teschner (1987) reports that the preparation of language department program directors, as evidenced by their dissertations, published writings, and presentations, is directed heavily to literature and not to the work of their actual positions. Such a finding is extremely important because it supports the supposition that language departments consider language program direction and research in language teaching and learning as unworthy of recognition for promotion and tenure. And yet, language program directors and teaching assistants could account for as much as two-thirds of the student credit hours generated in a language department.[2] There is certainly cause for concern in this situation.

3. Expectations for student language competence – Senior professorial staff are interested in literary and linguistic research, thereby effectively removing themselves from issues of language learning and teaching. They have relegated their responsibility to the "language coordinator." But, the expectations of this senior staff for language competence in the four modalities in upper division language courses is at the Superior level of the guidelines, creating tension between attention to language development and an ideal, but and unrealizable competence. The implementation of the guidelines in lower division courses makes that tension grow because some instructors will probably pay more attention to the development of language competence in students and less to outrageous expectations of senior faculty. This tension needs to be resolved if the guidelines are to be implemented at all.

4. Curriculum design – The lack of conceptualization and weak implementation of any curriculum design also hinder language learning, not only on the university level, but on the secondary school level as well. The lower division college curricula, like those in the secondary schools, are largely textbook oriented. Course syllabi reflect coverage of materials only in the text. Instructors seem to be driven to cover the syllabus. In other words, there is little attention paid to concepts of language development in language programs, but there is attention given to the learning of the "parts and pieces" of language. If the guidelines as a broad framework are to be implemented at the college/university level, attention must be drawn to the purposes and uses of language and its development in curriculum design. Crawford-Lange and Lange (1984), in their integration of language and culture, and Mohan (1986), who argues for the use of language as a medium of learning, point language learning away from almost total focus on the grammatical syllabus toward an integration of language with content. Stern (1983) in his statement of a multi-dimensional curriculum puts this direction in a broader focus. Such direction gives purpose to language learning.

5. **Language learning materials** on the university as well as on the school level, in spite of all the proficiency hype, are still focused largely on the grammatical syllabus. If the guidelines are to be implemented, materials must be focused more directly toward the use of language in the four modalities. While textbook authors may be in favor of this direction, publishers consider what sells. Their texts generally reflect a conservative response to "current" initiatives both because they are conservative and because it takes five years to write, field test, and produce a workable text. If the guidelines are to be implemented, then a curricular and instructional program will of necessity have to include the "practice of language use." Time will have to be devoted to oral situations performed in small groups, to listening and reading a variety of carefully chosen texts on topics of interest to students about the other culture(s), and to writing in the language for a purpose. Without such "practice," these uses may never be learned. The guidelines provide the framework for such a direction.

In spite of the many calls for the reorganization or reduction of grammar for communicative language teaching and learning (Brumfit, 1980; Valdman, 1980; Warriner, Gradisnick, Miller, Strasheim, & Valette, 1978), relatively little progress has been made in this regard. In a survey of state supervisors of foreign language, respondents indicated that teacher preference for the contents of language learning materials was heaviest on linguistic content in the four areas to which they were to respond: linguistic content, culture, communicative activities, global awareness. It is safe to assume that this situation is no different in higher education. The proficiency guidelines suggest that a balance among content, function, and accuracy be developed. To move away from the grammatical syllabus, some balance with other aspects of language and content is necessary.

6. **Class size** – With communicative outcomes for language learning taking prominence in both teachers' and students' minds, the importance of class size cannot be overestimated. If class size exceeds 25 students[3], and many times that is the case, the emphasis of the course will tend to focus on aspects which are easiest to teach and evaluate: grammar and translation. The practice of oral language use through situations and simulations becomes unwieldly and difficult to organize. The evaluation of oral language has similar problems. Listening comprehension, reading comprehension, and writing in language use activities require a different classroom organization. Large classes almost insure they will not be accomplished.

7. Current practice in classroom testing in university programs is a further inhibitor in the implementation of the guidelines. With a focus largely on grammar and vocabulary and what is covered in the text, it is largely associated with the written word. Little real attention is paid to listening and speaking in the evaluation process. If communicative outcomes are to become realizable, then students must be held accountable not only for grammar and vocabulary in the usual manner, but also for answering and asking questions orally, for the comprehension both oral and written texts never before heard or seen, and writing sentences or paragraphs about themselves.

If the proficiency guidelines and communicative outcomes are to be used in program development and serve as the framework discussed for curriculum, materials, and testing, then language programs at whatever level cannot continue "with business as usual." The preparation of teachers, specifically at the university level, must change. Programs must prepare teachers in an understanding of second language development in both children and adults and the interrelationship of that development with curriculum, instruction, materials, and testing in the classroom. Further, their functioning as teachers in the development of language competence must be regularly observed and evaluated. Class size must also be reduced. Without attention to these matters, the development of language competence for university students will remain at the level of "parts and pieces."

Articulation – The implementation of the ACTFL/ETS/ILR Guidelines may be useful in resolving the problems of articulation among and across educational levels by specifying learner outcomes at certain stages, communicating those outcomes broadly to students, parents, counselors, and administrators, and by evaluating those outcomes according to their descriptions or definitions (Lange, 1982). Here, the concept of the guidelines as a basic framework in which teachers and learners can function is important in providing continuity among educational levels (vertical articulation) and among courses designated for the same purposes on the same level (horizontal articulation). It is the experience of the Minnesota project that the development of tests forces consideration of articulation (the so-called "washback effect").

If the guidelines are to be implemented on the university level, great care must be taken to communicate expectations to secondary school programs. In fact, the Minnesota experience suggests that representatives from secondary schools and other colleges and universities in a region need to participate at almost every major level of decision-making. Such communication and participation reduces fears and suspicions. It also allows colleagues, working for the same purposes, to share common goals through a common language. The result is smoother transitions for students from secondary school to university programs. Without the descriptive statements of competency provided by the guidelines framework, however, communication across and between educational levels will remain difficult.

6. A final statement

At the University of Minnesota, an ambitious agenda has begun in the testing of student language competence in listening, reading, speaking and writing for both an Entrance Standard and a Graduation Requirement. The tests have been developed from the statements of language competence found in the ACTFL/ETS/ILR Guidelines. In such large scale testing, conscious compromises with reality have been made. We are using electro-mechanical means to evaluate listening and to serve as the source of a recorded oral interview. Listening and reading have been prepared as objectively scored tests. The writing test, like that of speaking, asks for production samples which are then judged by humans. The system is not perfect, but it has made a difference.

The implementation of the guidelines, particularly at the university level, presents some challenges to that system, namely preparation of teachers, a common direction for faculty, curricula and materials which develop language competence, and a testing program more oriented to language use. The guidelines themselves need constant and further work in depicting a complete a picture of language development and a less elitist nature. While not represented as a panacea, either here or in their use in the Minnesota program, the guidelines can be useful as a general framework of language competence for articulation across and among educational levels.

Notes

[1] The key wording here is proficiency; however, the mistaken notion of tying proficiency with number of quarters of study is to give a means of understanding the requirement to administrators, students, and the general public. The implementation of the policy, however, is functioning on the notion that seat-time does not necessarily guarantee competence in the language.

[2] This information comes from an unofficial accounting for student credit hours in the language departments at the University of Minnesota.

[3] Department chairs of French and Italian, German, and Spanish and Portuguese took a firm stand in AY 1986-87 to not admit more than 25 students to sections of these languages in first and second year.

References

American Council on the Teaching of Foreign Languages. (1982). *ACTFL provisional proficiency guidelines.* Hastings-on-Hudson, NY: Author.

American Council on the Teaching of Foreign Languages. (1987). ACTFL proficiency guidelines 1986. In H. Byrnes & M. Canale (Eds.), *Defining and developing proficiency: Guidelines, implementations and concepts* (pp. 15-24). Lincolnwood, IL: National Textbook.

Arendt, J. D., Lange, D. L., & Wakefield, R. (1986). Strengthening the language requirement at the University of Minnesota: An initial report. *Foreign Language Annals, 19,* 149-156.

Bernhardt, E. B. (1986). Reading in the foreign language. In B. H. Wing (Ed.), *Listening, reading, writing: Analysis and application* (pp. 93- 115). Middlebury, VT: The Northeast Conference on the Teaching of Foreign Languages.

Brown, C. L. (1984). The challenge for excellence in curriculum and materials development. *The challenge for excellence in foreign language education* (pp. 107-144). Reports of the Northeast Conference on the Teaching of Foreign Languages. Middlebury, VT: The Northeast Conference.

Brumfit, C. J. (1980). From defining to designing: Communicative specifications versus communicative methodology in foreign language teaching. *Studies in Second Language Acquisition*, 3, 1-9.

Crawford-Lange, L. M., & Lange, D. L. (1984). Doing the unthinkable in the second-language classroom: A process for the integration of language and culture. In T. V. Higgs (Ed.), *Teaching for proficiency, the organizing principle* (pp. 139-177). The ACTFL Foreign Language Education Series. Lincolnwood, IL: National Textbook.

Foreign language proficiency standards for entrance and graduation, College of Liberal Arts: A guide for teachers of French. (1986). Minneapolis: College of Liberal Arts, University of Minnesota.

Foreign language proficiency standards for entrance and graduation, College of Liberal Arts: A guide for teachers of German. (1986). Minneapolis: College of Liberal Arts, University of Minnesota.

Foreign language proficiency standards for entrance and graduation, College of Liberal Arts: A guide for teachers of Spanish. (1986). Minneapolis: College of Liberal Arts, University of Minnesota.

Galloway, V. (1987). From defining to developing proficiency: A look at the decisions. In H. Byrnes & M. Canale (Eds), *Defining and developing proficiency: Guidelines, implementations and concepts* (pp. 25- 72). The ACTFL Foreign Language Education Series. Lincolnwood, IL: National Textbook.

Heilenman, K. L., & Kaplan, I. M. (1985). Proficiency in practice: The foreign language curriculum. In C. J. James (Ed.), *Foreign language proficiency in the classroom and beyond* (pp. 55-78). The ACTFL Foreign Language Education Series. Lincolnwood, IL: National Textbook.

Lange, D. L., & Lowe, P., Jr. (1986). Testing the internalizability of the ACTFL/ETS/ILR reading proficiency scales. *Paper presented at the Language Testing Colloquium*, Monterey, CA.

Lange, D. L. (1982). The problem of articulation. In T. V. Higgs (Ed.), *Curriculum, competence, and the foreign language teacher* (pp. 113-137). The ACTFL Foreign Language Education Series. Lincolnwood, IL: National Textbook.

Mohan, B. A. (1986). *Language and content*. Reading, MA: Addison-Wesley.

Stern, H. H. (1983). Toward a multidimensional foreign language curriculum. In Mead, R. G., Jr. (Ed.), *Foreign languages: Key links in the chain of learning* (pp. 120-146). Reports of the Northeast Conference on the Teaching of Foreign Languages. Middlebury, VT: The Northeast Conference.

Teschner, R. V. (1987). A profile of the specialization and expertise of lower division foreign language program directors in American universities. *Modern Language Journal, 71*, 28-35.

Valdman, A. (1980). Communicative ability and syllabus design for global foreign language courses. *Studies in Second Language Acquisition, 3*, 81-96.

Warriner, H. P., Gradisnick, A., Miller, J., Strasheim, L. A., & Valette, R. (1978). High school foreign language texts: Too much between the covers to cover. *Foreign Language Annals, 11*, 551-557.

Appendix

University of Minnesota, College of Liberal Arts, Foreign Language Proficiency Tests

Listening

The listening test is 40 items long. Students will complete the test in approximately 30 minutes, depending on the language. Each question is related to a "text" that students hear before they answer questions on it. Depending on the level of the test, the "texts" are repeated. And also depending on level, there will be more than one question per "text." The questions are multiple choice in nature. The answers will be machine scored. The texts have been developed from authentic materials.

Reading

The reading test is 40 items long. Students will be given 30 minutes to complete the test. Each question is related to a text that students will read before they answer questions. In many instances, there will be more than one question per text. A response to the questions comes through multiple-choice answers. The texts are from newspapers, magazines, advertisements, and other texts which have been chosen and graded in difficulty for this test. The test is machine-scored.

Writing

The length of the writing test varies by language. Students will have 30 minutes to complete the test. Students will write answers on the tests themselves. They will complete such tasks as writing lists, notes, short compositions, and the like. The responses will be scored according to three criteria: content, function, and accuracy and will answer the following questions: Do students make the kinds of error expected at this level (fewer or more) and are students able to do the task asked of them (write a list, notes, a short composition, and the like)? The test is scored using a rating system for each item; it will be rated by trained raters.

Speaking

The speaking test is about 10 minutes in length. It is in the form of an interview. The questions require answers related to the person taking the test. They are related to personal information, situations in which students are placed, descriptions, narrations and the like. The responses are judged according to three basic criteria: content, function, and accuracy. The system for making those judgments is similar to the criteria for judging competence in writing.

REACTION TO LANGE AND FREED PAPERS

Robert C. Lafayette
Louisiana State University

"From the perspective of many language faculty at the University of Pennsylvania, the commitment to proficiency-based instruction has improved the language learning environment at Pennsylvania. There is no question that the investment in instructor energy in terms of training, teaching and testing has been enormous, but the results seem to make it worthwile." (Freed, I.U. Symposium, March, 1987). Immediately following this statement, Freed goes on to say that they are still unable to demonstrate that proficiency-based instruction actually improves language instruction. Such may well be the case, but what Freed fails to say on her own behalf is that although they have no evidence of improved language learning, their adoption of proficiency-based exit criteria for their foreign language general education requirement at least provides evidence that those students who have successfully completed the requirement have indeed achieved the goal set by the university. The University of Pennsylvania knows what their foreign language students have accomplished; similar information would be virtually impossible to obtain in most other university programs.

Lange and Freed have to be congratulated for their enormous efforts, for their vision and for their courage. What they have done or are about to do is so basic to the principles of education that it defies elaboration. They have (1) set their goals, (2) selected and/or designed materials to meet those goals, and (3) chosen or designed appropriate instruments to test these goals. Unfortunately, too many courses lack this basic internal articulation. What is found instead are either courses without goals, courses with inappropriate materials or courses where the final exam reflects little of what was taught in class. In his paper Lange states the following: "At the University of Minnesota we recognize the arguments both pro and con concerning the nature of guidelines as well as their several flaws. However, we make no apologies for using the ACTFL Guidelines: we decided to use them in spite of the flaws. The ACTFL Guidelines are the only complete system upon which we could build our requirement. The decision to use these guidelines was a conscious one; we did it with open eyes."

I, too, make no apologies whatsoever for using the ACTFL Guidelines in teacher certification, the area upon which the guidelines and the Oral Proficiency Interview, in my opinion, can have their most significant impact. It is my firm belief, one which I hope is shared by many, that a foreign language course ought to be taught primarily in the foreign language. However, in order for that to be the case, the teacher must know more than a set of grammar rules. He/she must have at least a basic level of oral proficiency. Unfortunately, such is not the case for a large proportion of both the teachers already in the field and those newly certified every year. 1984 pilot test results of OPI interviews of Texas foreign language education majors show that 52.3% of those students tested were rated Advanced or better on the ACTFL scale. The rest were at Intermediate High and below, with some even in the Novice range.[1] Unfortunately, this problem is not unique to Texas. This summer in Louisiana we are hosting a six week institute for Louisiana teachers of French who also are at the 1+ Level.

One of my primary concerns as a teacher trainer is to see that our candidates have the language skills necessary to teach. The Guidelines and the OPI serve that concern well. They permit us as a profession to be held accountable. As Lange said, they are the only complete system upon which we can set our goals and test them. And what is more important, it works.

In early 1984, I made a personal decision that I needed to know more about the Oral Proficiency Interview and the ACTFL Guidelines. I therefore applied to the Illinois Foreign Language Teachers Association who had a grant to train individuals from throughout the midwest free of charge. I spent five days in one of the most professionally run workshops I had ever attended and upon my return I saw the enormous potential of the instrument in the area of teacher certification. I was subsequently instrumental in establishing exit requirements for Indiana University's teacher training candidates which were added to the already existing State regulations. At the time the State of Indiana had seat time regulations. We superimposed our own to improve our product. In the area of oral proficiency the regulation stated that students had to achieve a level of 2+ (advanced-plus) prior to student teaching. Consequences of this new regulation were (1) increased study abroad, (2) students seeking courses in which they knew they would have an opportunity to converse, (3) students exchanging conversational time with foreign students learning ESL, and in some cases, (4) students complaining about their 40 plus credits of foreign language courses which did not prepare them to pass the oral test.

A similar regulation exists at Louisiana State University, with the difference that in Louisiana, the State mandates the National Teacher Examination including the specific foreign language test; the state also mandates study abroad for French candidates and in some cases it mandates the OPI. In those latter cases, testers must be officially certified by ACTFL and have a copy of that certificate in the State Teacher Certification Bureau.

The OPI is also used in Louisiana for the selection and classification of candidates for its extensive summer abroad program. During the spring of 1987, oral interviews were administered to more than 100 candidates for summer scholarships abroad. In order to create more homogeneous groups, students and in-service teachers were assigned to one of five locations abroad, based partly on OPI scores. During a 1987 six-week summer institute held both in Baton Rouge and in La Rochelle, France and sponsored jointly by the French government, the Louisiana Endowment for the Humanities and Louisiana State University, the OPI was used to select and evaluate the twenty participating teachers. One of the Institute's goals is to improve the oral proficiency of teachers at or near the Intermediate High Level. During the first week of the program, two certified testers not only determined the proficiency level of each individual but also wrote an individual oral language prescription for each participant. This prescription then provided the basis for determining the oral program during the four-week stay in La Rochelle. Finally, the OPI was administered once again during the last week of the program in Baton Rouge. Various data were gathered throughout the Institute and will be made available in the near future.

The most often voiced criticism of the Oral Proficiency Interview is that it lacks validity. Although such may be the case, it serves preservice and inservice teacher training programs very well and most importantly it encourages teacher candidates to meet minimum standards. There are currently some twenty states studying the various possibilities of using the OPI for certification. Such action can only improve the proficiency of future teachers. Similarly, ACTFL's Teacher Education Standards Commission is presently considering recommending that teachers be proficient at a level equivalent to Advancd Plus in listening, reading and speaking, and at a level equivalent to

Advanced in writing.

In conclusion, it must be stated that the ACTFL Guidelines and the Oral Proficiency Interview have had and continue to have a great deal of impact on the teaching of foreign languages in the United States. In fact, this symposium would not have been held were it not the case. For the most part, the impact has been extremely positive due to the fact that it has encouraged numerous states and institutions to take stock of their teachers and students. If the Guidelines and the OPI are partially flawed, steps should be taken, not to destroy them, but rather to repair them, for too much good has come, is coming and will come from what already exists.

Note

[1] Hiple, D. V., & J. H. Manley. (1987). Testing how well foreign language teachers speak: A state mandate. *Foreign Language Annals*, *20*, 147-53.

DISCUSSANT PAPER ON THE IMPLEMENTATION OF THE ILR/ACTFL GUIDELINES AT THE HIGH SCHOOL AND UNIVERSITY LEVELS

Sandra Savignon
University of Illinois at Urbana-Champaign

Barbara Freed and Dale Lange have given accounts of initiatives in the evaluation of second/foreign language abilities at two major U.S. universities. The examples they have given are those of responsible, thinking educators who have adapted the ACTFL guidelines to suit their own contexts of teaching. I know that, in each case, this has required great personal commitment of time and talent, and I feel we should applaud them on their efforts. In fact, I was happy to include Barbara Freed's account of the implementation of a proficiency requirement at the University of Pennsylvania in a collection of initiatives in communicative language teaching that I edited a few years ago with Margie Berns (Savignon & Berns, 1984).

There were other initiatives in that same collection. They included an account by Gerngross and Puchta (1984) of the development of communicative teaching materials for learners of English in Germany, a paper they subtitled "The Art of Letting Go." There was also a report by Sauli Takala (1984) on the curriculum reform in Finland, a reform that, he cautioned, required fully 10 years to complete. It involved interaction between administration, government and teachers at all stages of development. There was also a report by Merrill Swain (1984) on the tests of communicative language use that are being developed for French immersion programs in Ontario, an initiative that ties in with the report given by Sharon Lapkin and Susanne Carroll earlier in this conference on "The Role of Vocabulary in the Development of Communicative Ability." And there was also a report by Anderson and Rhodes (1984) on immersion programs in U.S. elementary schools.

What, then, is the problem with proficiency? Or as Bob Lafayette quipped the other day, with respect to the issue of replacing a letter grade evaluation of learner achievement with a number on an ILR/ACTFL proficiency scale, "What's the big deal?" The big deal, or the problems, as I see them, are the following: (1) First of all, the ILR/ACTFL guidelines offer an outdated model of language behavior. It is a "four skills plus culture" model clearly reminiscent of audiolingual thinking. (2) Second, the guidelines make the unfounded claim that they represent stages of language proficiency "regardless of where, when, or how [it] has been learned or acquired" (ACTFL, 1986). (3) And third, the guidelines originated in government diplomatic and military agencies and are not based on an assessment of learner needs in U.S. school and university programs. (See Savignon, 1985).

This would be all right if we could ignore them. But a national professional organization, the American Council on the Teaching of Foreign Languages, has developed and promoted them, with considerable help, to be sure, from military and other government testers. While there have been disclaimers from some of those involved with guideline development as to their curricular and methodological implications, others have not been at all shy about making claims for teaching. Earlier today, Pardee Lowe held up the spectre of the "Terminal 2 profile" and places the blame on teacher failure to correct errors---that is, deviations from some adult native norm---in learner classroom oral expression. These undocumented claims (see Savignon, 1987) figure prominently in the methods book that has been written by the immediate past president of ACTFL, a person intimately involved with both writing the guidelines and promoting their use (Omaggio, 1986).

In what purports to be a guide to "proficiency-oriented instruction," there is no critical examination of current theories of language learning both inside and outside the classroom. Rather, Krashen's (1982) hypotheses regarding natural language acquisition are summarily dismissed, and an early research project on the development of communicative competence in a classroom setting (Savignon, 1972) is erroneously represented as encouraging learners "to abandon the linguistic system per se" in their attempt to express meaning (from Higgs & Clifford, 1982, p. 4). While acknowledging the negotiative nature of communication, Omaggio again defers to Higgs and Clifford in pointing out that "this view of communicative competence does not lend itself easily...to determining in any precise way the degree to which an individual is communicatively competent" (p. 5).

Proficiency, as defined by the ILR/ACTFL Guidelines, presumably provides the needed precision: "The most recent conceptualization of the term proficiency, includes specifications about the levels of competence attained in terms of the functions performed, the contexts in which the language user can function, and the accuracy with which the language is used." (p. 8). Omaggio subsequently claims support in non-existent Higgs and Clifford data for insisting on grammatical accuracy in the beginning stages of learner oral expression and provides examples of the "contextualization" of surface-level grammatical features. Texts are used as pretexts in a structurally-driven, as opposed to a meaning-based, syllabus. In sum, the teaching methods presented are little more than ALM regarbed.

Where do we go from here? I suggest that for the moment we say Basta! and that we suspend revision and promotion of the Guidelines. The last thing that I would like to see emerge from this conference is a set of "1987 Revised Provisional Tripartite ACTFL Guidelines." In the meantime, I think there needs to be an informal dialogue within ACTFL, and within the academic community, along with small scale exploration of learners' needs and interests in various U.S. contexts. If a national curriculum is to be proposed with corresponding achievement tests, my hope is that it would reflect current understanding of language and language behavior, the kind of understanding that Ron Wardhaugh (1985) presents so nicely in his book, How Conversation Works. Or the model that Canale and Swain (1980) and Savignon (1983) have offered as classroom models of communicative competence. Communicative language teaching involves communication, which in turn involves a community with shared assumptions and a willingness to negotiate meaning. A plan to improve the second/foreign language communicative competence of learners in U.S. schools and universities should take the view not that we should teach what we can test, but that we will find ways to measure that which we understand to be important.

References

American Council on the Teaching of Foreign Languages. (1986). *ACTFL proficiency guidelines*. Hastings-on-Hudson, NY: Author.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second-language teaching and testing. *Applied Linguistics*, 1-47.

Freed, B. (1984). Proficiency in context: The Pennsylvania experience. In S. Savignon & M. Berns (Eds.), *Initiatives in communicative language learning* (pp. 211-240). Reading, MA: Addison-Wesley.

Gerngross, G., & Puchta, H. (1984). Beyond notions and functions: Language teaching or the art of letting go. In S. Savignon & M. Berns (Eds.), *Initiatives in communicative language learning* (pp. 89-107). Reading, MA: Addison-Wesley.

Higgs, T., & Clifford, R., (1982). The push toward communication. In T. Higgs (Ed.), *Curriculum, competence, and the foreign language teacher*. ACTFL Foreign Language Education Series, vol. 13 (pp. 57-59). Lincolnwood, IL: National Textbook Co.

Krashen, S. (1982). *Principles and practice in second language acquisition*. New York: Pergamon Press.

Omaggio, A. (1986). *Teaching language in context: Proficiency-oriented instruction*. Boston, MA: Heinle and Heinle.

Savignon, S. (1972). *Communicative competence: An experiment in foreign language teaching*. Philadelphia: Center for Curriculum Development.

Savignon, S. (1983). *Communicative competence: Theory and classroom practice*. Reading, MA: Addison-Wesley.

Savignon, S. (1985). Evaluation of communicative competence: The ACTFL provisional guidelines. *Modern Language Journal*, *69*, 129-134.

Savignon, S. (1987). Second-language acquisition/foreign language learning: Nothing is more practical than a good theory. *Paper presented at the Second-Language Acquisition/Foreign Language Learning Conference*. Urbana-Champaign: University of Illinois.

Savignon, S., & Berns, M. (1984). *Initiatives in communicative language learning*. Reading, MA: Addison-Wesley.

Swain, M. (1984). Large-scale communicative language testing: A case study. In S. Savignon & M. Berns (Eds.), *Initiatives in communicative language learning* (pp. 185-201). Reading, MA: Addison-Wesley.

Takala, S. (1984). Contextual considerations in communicative language teaching. In S. Savignon & M. Berns (Eds.), *Initiatives in communicative language learning* (pp. 23-24). Reading, MA: Addison-Wesley.

Wardhaugh, R. (1985). *How conversation works*. New York: Basil Blackwell.

COMMENTS ON THE FREED AND LANGE PAPERS

Protase Woodford
Educational Testing Service

Both Freed and Lange highlight in their papers the need for compromise in the development of operational testing programs at their institutions. At both universities it was important to respond to the needs and desires of a variety of potential "stakeholders" in order to enlist their cooperation or, at least, to avoid their opposition.

These compromises resulted in assessment instruments and descriptions less than ideal. For example, the Language Advisory Committee of the University of Pennsylvania developed the following statement of purpose:

The language requirement should prepare students to use a foreign language meaningfully. This includes the skills that are required to participate at a minimal level of proficiency in authentic target-culture situations that might call for listening, speaking, reading and/or writing. Such skills will, in principle, span a continuum from such day-to-day tasks as giving directions to those skills required in analyzing a work of art. This required level of competence is that which would be minimally acceptable to native speakers and can reasonably be acquired by students in the time normally devoted to required language study. At this level of minimal proficiency, importance is given to grammatical accuracy and the ability to express one's intent in a variety of situations. To those familiar with the ILR and ACTFL scales, a lack of concordance is apparent between this "Requirement" statement and the "Minimum passing proficiency levels". Moreover, it is questionable whether an intermediate level speaker or writer can "analyze a work of art" through the medium of the foreign language, and the "importance given to grammatical accuracy" seems at variance with the ACTFL descriptions of an intermediate level writer or speaker:

Intermediate:	The Intermediate level is characterized by the speaker's ability to: -create with the language by combining and recombining learned elements, though primarily in a reactive mode; -initiate, minimally sustain, and close in a simple way basic communicative tasks; and (-ask and answer questions.)
Intermediate-High:	Able to meet most practical writing needs and limited social demands. Can take notes in some detail on familiar topics and respond in writing to personal questions. Can write simple letters, brief synopses and paraphrases, summaries of biographical data, work and school experience. In those languages relying primarily on content words and time expressions to express time, tense, or aspect, some precision is displayed; where tense and/or aspect is expressed through verbal inflection, forms are produced rather consistently, but not always accurately. An ability to describe and narrate in paragraphs is emerging. Rarely uses basic cohesive elements, such as pronominal substitutions or synonyms in written discourse. Writing, though faulty, is generally comprehensible to natives used to the writing of nonnatives.

These apparent inconsistencies are cited not to criticize, but rather to emphasize the constant need for compromise. How do you move colleagues who view a focus on functional language skills as "anti-humanistic", "anti-intellectual" and "Berlitz-like" to acceptance of a "proficiency based foreign language requirement"? Including reference to "skills required in analyzing a work of art" and "importance...given to grammatical accuracy" may help.

The situation confronting Lange and his colleagues at Minnesota was somewhat different from that faced by Freed and her colleagues. Whereas at the University of Pennsylvania the key issue seemed to be how to bring about agreement among independent language departments and faculty members on the need for a proficiency requirement and an appropriate evaluation process; at Minnesota a task force presented recommendations for such a requirement which were accepted without dissent. It is noteworthy that at Minnesota as at Penn a link with "seat time" was still evident in official statements, i.e., proficiency was characterized as that "usually attained after three quarters of college study"; "proficiency...usually attained after six quarters of study."

At Minnesota a large scale testing program for yearly populations in the thousands was to be designed and implemented. The Minnesota project had to deal with the classic problems facing all large scale testing programs: the need to demonstrate reliability and validity; to establish a scoring and reporting system; to be cost effective and efficient.

In order to "deliver" the testing program, another series of compromises between the ideal and the "do-able" had to be made. For example, contrived stimuli for the listening tests were used instead of real-life speech samples. A "semi-direct" (as John Clark would call it) or "canned" speaking test was substituted for the oral proficiency interview.

In 1986 the Connecticut State Department of Education contracted with ETS for the development of Listening, Reading, Culture and Writing tests in French, German, Italian and Spanish. A battery of Latin tests was also part of the project. The modern language tests were to be developed at two levels; Intermediate, for students at the end of two years of foreign language instruction; and Advanced, for students at the end of three or more years of instruction. The test specifications required that the tests and each test question be identified by an "ACTFL Level."

Panels of experienced teachers familiar with the ACTFL scale were requested to assign a level to each test question. Connecticut Department of Education Testing Specialists and ETS staff also labeled the questions. Over 500 test questions for Listening and Reading were assigned levels by 5 to 7 people. I recall not one case in which there was unanimous agreement on a level, nor a case in which a single panel of teachers was in total agreement. So disparate were the judgments that it was decided to collapse the novice low-mid-high distinctions into a single "novice" level, and the intermediate low and mid into a single "intermediate level." It is ironic that this decision was almost the reverse of the original decision to burst out the ILR S-O and S-1 levels into the three ACTFL levels for each.

In order to provide listening and reading tests for mass administration a machine-scorable format was required. In Connecticut, as in Minnesota, a multiple choice format was chosen. A "pure" listening test should not require the examinee to read in the foreign language. Some multiple choice formats that are "pure" are spoken stimulus plus illustrations as options, spoken options, or options printed in English. There are advantages and drawbacks to each format. Illustrations are expensive. Spoken options introduce a memory factor, especially with four choices. Printed English options are anathema to many foreign language teachers.

The final format of the listening test was the result of a number of compromises such as the use of three spoken options instead of four to reduce the memory load resulting in a possible reduction in item reliability, and the use of printed foreign language options instead of English to repond to strong teacher objections to a mixed-mode format.

Any major testing effort beyond the level of an individual classroom is, necessarily, the product of compromise, as we are reminded by Freed and Lange. Faculty must be stakeholders. Their needs and opinions must be heard and be reflected in test specifications. A pschometrically elegant test that is not accepted by the language teaching profession is of little value. As appealing as the ACTFL scale and guidelines are, it is nonetheless unlikely that they will totally replace traditional test statistics in the evaluation of foreign language skills. What is more likely is the use of some combination of ACTFL descriptions and indices of difficulty and reliability such as p values and biserial correlations. In our work for Connecticut, in connection with pre-testing, we found many cases where test exercises considered "novice" on the ACTFL scale proved difficult for the sample population while "intermediate" exercises were easier than those labeled "novice." The availability of traditional data will allow teachers and other decision makers to see how well the outcomes of instruction accord with the various ACTFL Levels and whether or not the "guidelines" are appropriate for their programs. Such data may also prove valuable as an aid in the refinement or revision of the scales/ guidelines.

The oral proficiency interview and scales have a history of some thirty years. The scales for listening, reading, and writing do not. The interview has been used with tens of thousands of examinees: government personnel, Peace Corps volunteers, bilingual and foreign language teacher candidates, missionaries, business people, high school and university students. As the scales for listening, reading and writing undergo extensive in-field application it can be expected that a need for further revisions will become obvious.

Lange indicates that such guidelines "may be useful in resolving the problems of articulation across and among educational levels." The potential of a language proficiency scale for aiding in resolving the problem of articulation was a motivating factor behind the original "Common Yardstick" effort. At the 1973 Georgetown University Roundtable, Woodford emphasized that in order to effect smooth articulation, what is needed is a widely recognized and accepted 'yardstick' by which to measure performance in a foreign language. He suggested that objectives be considered within two categories. The first would contain detailed, specific objectives to be useful primarily to the teacher for more effective instruction and the second, more general performance objectives that could be used to define levels of competence. As an example of the second (general performance objectives) category the Foreign Service Institute Scale for Levels of Speaking Proficiency was cited. For such a scale or set of performance criteria to have maximum value and applicability, all potential users would need to be thoroughly familiar with the performance represented by each step on the scale.

A number of Symposium papers caution against negative results of over use or misuse of the guidelines. Lange sees teachers as "tired of bandwagons" and considering the "Proficiency Movement" as just another one. The "Proficiency Movement" has moved far beyond the modest goal of its beginning: a set of descriptors to illustrate different levels of speaking ability that would be of help to language learners and teachers. The government scale was insensitive at the lower end where most school and college learners would find themselves. A refinement of the lower end providing more descriptions was undertaken. From that beginning we now have training programs, "certified testers", curriculum workshops, methodologies, texts, all carrying a "proficiency" label. Lange is justified in warning us of teachers being tired of bandwagons and becoming bored with "proficiency".

Scales for speaking have proved useful in a number of contexts. Scales for other skills could also be valuable. The scales are not scientifically determined nor necessarily accurate. They are not perfectly suited to all applications, either. For those who understand the scales and their limitations, the scales are helpful. Unfortunately an adverse reaction toward the "proficiency movement" could result in the disuse of a useful tool.

Reference

Woodford, P. et al. (1973). Coordination of high school and college objectives in foreign language instruction. In K. R. Jankowsky (Ed.), *Georgetown University Round Table on Languages and Linguistics (GURT)*, 231-238.

Summary of the discussion session on PRACTICAL PROBLEMS IN THE ADMINISTRATION OF THE ILR-ACTFL PROFICIENCY TEST AND THE IMPLEMENTATION OF THE GUIDELINES AT THE HIGH SCHOOL AND UNIVERSITY LEVELS

Patricia J. Kyle and Laureen Hurt
Indiana University

Comments centered around two topics, each discussed in some detail: (1) the Oral Proficiency Interview (OPI) as an assessment instrument and (2) the ILR-ACTFL Proficiency Guidelines and OPI as models for language instruction.

1. The proficiency test as an assessment instrument

1.1. The OPI has been used for program evaluation – This is to say that some programs are being evaluated on the basis of how well they prepare learners to meet proficiency guidelines. In response to P. Woodford's contention that the guidelines have not stimulated a great deal of change in the language classroom, it was pointed out that the ACTFL proficiency guidelines influence educational decisions made at a level above that of the classroom teacher. The teacher makes decisions concerning instruction and testing in the classroom where the impact of the guidelines may be minimal. However, it was argued that decisions about program evaluation, teacher certification and proposals are being made with reference to a set of criteria presented in the guidelines. Therefore, the overall impact of the guidelines is not at all minimal. One participant argued that there is a great need for program evaluation in the field of foreign language teaching because the profession has a history of accepting methodologies on the faith of testimonials given by their practitioners. The participant maintained that the statements by Freed and Lange are examples of such testimonials and that we as a profession can no longer continue accepting these various approaches on faith alone. What is different about the programs described by Lange and Freed, according to the participant, is that they are well defined enough that perhaps it is possible to systematically evaluate them. However, he did not think that the criteria used in the OPI should be used for program evaluation. Another participant remarked that we need to know more about the actual content of the programs at the University of Minnesota and the University of Pennsylvania and not just about how the end results are assessed. Barbara Freed maintained that even though the University of Pennsylvania program was already well established and quite sophisticated when the ACTFL proficiency guidelines were adapted for use there, its achievements could be replicated. Freed stressed the importance of cooperation among members of the language teaching profession to resolve the controversies concerning the OPI and guidelines. She warned against discarding them completely and thus allowing proficiency to become another "bandwagon".

1.2. The OPI has been used as a component in the teacher certification process in some states – Some participants addressed the question of the statistical validity of the OPI and its suitability for use in teacher certification. One participant responded to Woodford's statement that psychometricians who are not language people, who often do not know the content of tests, and who are mainly occupied with numbers, also have some say in deciding what tests are implemented in a particular state. This participant pointed out that there are psychometricians who are interested in test item validity. Woodford asserted that, for teacher certification, he would rather rely on a speech sample drawn using the OPI than the NTE (National Teachers' Examination) which does not test the teacher's ability to speak the foreign language.

1.3. The OPI has been used for evaluating language learners – In reaction to Savignon's criticism that the OPI tests without regard for where or how language was learned, a participant pointed out that it is not necessary to take into account the way in which language was learned in order to be able to assess it. Further, in reaction to Savignon's statement that the OPI does not take into account learner needs, the same participant noted that it is also not necessary to consider learner needs in order to assess language proficiency.

2. The OPI and ILR-ACTFL Guidelines as models for classroom instruction

2.1. The guidelines have been used in developing teaching practices – Savignon asserted that learner needs, and not a test like the OPI, should determine classroom instructional goals. A participant responded to this by arguing that we cannot know what learners' needs are, that the needs of the learners evolve with their lives. The participant went on to specify that we cannot know what learner needs are on more than a local basis and that this is not sufficient if we want to be held accountable to the outside world for the language training we give. Another participant argued that there is no point in doing anything if the need for doing it is not known. The matter of who determines learner needs was brought up when another participant called attention to the fact that college freshmen cannot articulate their language learning needs. Still another participant pointed out that, as the educators, we are the ones who articulate those needs for them. Savignon maintained that the teachers, learners, and learners' parents must all be involved in the statement of needs.

2.2. The guidelines have been perceived by some to reflect the language learning process – Savignon's criticism that the OPI and ACTFL guidelines claim to represent stages of acquisition was met with some disagreement. One participant pointed out that the OPI and guidelines never claimed to reflect stages of language learning, that this is still in the process of being developed, and that it is a long way from being worked out. Another participant suggested that the problem may center around short term as opposed to long term solutions. He maintained that the guidelines are not adequate for intellectually educating teachers as to what language acquisition is. Research has been trying to answer the question of what language acquisition is for 15 years and will keep on trying to answer it for the next 1500 years. He warned that there is a danger in selecting a test that offers only short term solutions because we risk producing a generation of teachers who do not think about what they are doing. Another participant responded to Savignon's claim that the OPI and guidelines are based on an outdated model of language. She pointed out that the model of language upon which the guidelines are based is still being developed. Savignon asserted that if the OPI is based on a model that is still being developed, then it is premature to use it for assessing learners, determining teacher certification, and selecting criteria to be used in program evaluation.

3. Conclusion

Questions raised in the discussion summarized above related to the topic of assessment as well as that of language instruction. They focused on the following areas of concern:

(1) The extent to which the OPI as an assessment instrument implicitly reflects a model of language learning.

(2) The impact at the various levels of language education (administrative, instructional, and learning) of the OPI and the Proficiency Guidelines as assessment instruments.

(3) The suitability of the guidelines as a framework for developing instructional practices.

(4) Finally, implicit in the overall discussion was the question: Is use of the OPI and ILR-ACTFL Proficiency Guidelines the best way to proceed in meeting assessment and curricular needs in an academic context?

EVALUATION OF FOREIGN LANGUAGE PROFICIENCY: SYNTHESIS

G. Richard Tucker
Center for Applied Linguistics

In the remarks to follow, I shall identify three threads which seemed to me to underlie much of our discussion at this symposium:

1. The need to develop shared operational definitions for a substantial number of key terms which would serve to enhance communication among the diverse individuals, from so many different disciplinary backgrounds, interested in foreign language proficiency;
2. The need to develop and implement a vigorous proactive research agenda; and
3. The need to undertake a broadly-based information dissemination campaign.

Background

By way of background, I should note that the needs to develop economical, sensitive, reliable and valid measurement instruments for all aspects of foreign language proficiency are at least threefold: for diagnosis and placement purposes, for certification or for exit criteria; and to provide formative feedback for program evaluation and modification. The "clients" to be served are quite broad. Indeed, part of the breakdown in communication at various points during the symposium was attributable to the fact that the Interagency Language Roundtable (ILR)/Educational Testing Service (ETS)/American Council on the Teaching of Foreign Languages (ACTFL) approach to foreign language proficiency tries to be all things for all audiences. The audiences, however, are extermely diverse -- government clients who require demonstrated high levels of proficiency in a particular language to be able to function in sensitive diplomatic missions via that language; academic audiences comprising of students who need to satisfy distributional requirements or wish to spend a portion of their academic program or summers living abroad on the one hand, or who are candidates for foreign language and area studies programs overseas.. The needs of secondary school clients are correspondingly different although not necessarily lower (e.g., consider the students who have participated in foreign language immersion programs), while the needs of the private sector were not addressed at this meeting at all despite the fact that a number of major multinational corporations have or are in the process of implementing common language policies for their worldwide offices and have set some type of language proficiency requirement.

Terminological differences

With these remarks by way of background, let me address the first thread which permeated our discussions. The meeting was marked, for me at least, by a series of dichotomies -- that is, by the use of a number of crucial or pivotal terms or concepts which seemed to recur prominently in our discussions and which were either: (1) ill-defined, or (2) were identified as key constructs which required further elaboration, definition, or research and to which we vowed to return but never did. These key dichotomies, in random order, follow:

generic vs. language specific guidelines
accuracy vs. appropriateness
the ability or trait to be measured vs. the method of measurement
western vs. non-western languages (and the consequences for proficiency assessment)
productive vs. receptive skills
needs of government clients vs. needs of academic clients
standardization with white collar clients vs. blue collar clients
codified or (standardized) languages vs. less well codified or standardized languages
the "high" form vs. the "low" form of a language as the standard
vertical integration vs. horizontal integration
a single global measure vs. component measures of proficiency
user judgment vs. tester/interviewer/rater judgment
rating as an end in itself vs. provision of feedback to the student
research vs. quantitative research
North American developments vs. (lack of knowledge about) European developments
language to communicate vs. language to mark ethnic identity
ACTFL guidelines for proficiency assessment vs. curriculum development
learning vs. acquisition
nonformal acquisition vs. formal training
top down models vs. bottom up models
product vs. process
Chomskyian vs. Firthian traditions

I have identified, without editorial comment, a number of the threads which seemed to recur during the symposium. I, at least, had the pervasive feeling that people were passionately committed to one or the other construct or viewpoint. And it often seemed as though people did not share an operational definition of terms such as accuracy, fluency, or even proficiency. My feeling is that a good deal of work must be done to ensure that the key concepts or underpinnings of the so-called proficiency movement are clarified and disseminated.

Proactive research agenda

I see a second need as the development a proactive and vigorous research agenda to accompany the elaboration and continuing use of the ILR/ETS/ACTFL approach to foreign language proficiency assessment. Let me allude for a moment to the model of research which accompanies the Test of English as a Foreign Language (TOEFL) program. There is a vigorous, directed, peer-evaluated research program which supports TOEFL test development and analysis efforts. Specific funds (a percentage set aside from income derived from testing) are available to permit the "commissioning" of both applied and basic research to describe test taker or population characteristics; to ensure the reliability of the instrument; to continually assess the validity of the instrument and its derivatives; and to assess a variety of factors associated with topics such as methods of delivery, etc. A specific and proactive long term agenda has been developed and research is suggested and evaluated with reference to its contribution to improving the quality of the test and to advancing the field of measurement. I suggest that those concerned with the

ILR/ETS/ACTFL approach to foreign language proficiency assessment consider, in similar fashion, the development of a long-term research agenda, specific mechanisms for commissioning and funding relevant research, and procedures to ensure the incorporation of information from that research in a formative manner into test revision. Let me now suggest several potential exemplary research activities that were discussed, at various points, during the symposium.

1. Secondary analysis of existing data should be conducted. In particular, a qualitative "content" analysis of audio and videotapes should be carried out. There exists a wealth of data which can and should be further analyzed.

2. Secondary analysis could be conducted using data from the CIA/DLI/FSI "comparability study" conducted by John Clark. Specifically, university raters could be asked to evaluate the same tapes that were previously evaluated by representatives of the three federal agencies. Or one could compare the ratings of representatives from University One with those from University Two. Further additional comparability studies might be conducted to examine within-rater reliability, between-rater reliability, or the reliability for bilingual judges versus monolingual judges across languages. In addition, attention should be paid to the distinction made by John Clark between "absolute" and "linear" reliability.

3. "Use-validity" studies should be carried out. What can a person do (with the language) when functioning in job X? What is the range of the person's abilities? What can the person not do effectively using the target language? Such studies should be done across a variety of occupations, in various in educational settings, and in various technical training programs.

4. We talked a good deal about "user" reactions, and I believe it would be appropriate to implement a series of "matched guise" studies to examine what factor or factors (phonological, lexical, syntactic, discourse) that trigger stereotypic reactions. We discussed the attitudinal reactions that might be elicited by a highly proficient individual and wondered whether at some point (cf., Einar Haugen's "The Ecology of Language") a speaker/learner would be motivated to maximize communication or to mark oneself as a member of a particular ethnic group. During the meeting, the example of the American in China arose and we wondered whether the individual would consciously or subconsciously think it advantageous to be perceived as a fluent speaker of Chinese -- but yet one who was an American--or whether the goal would be to present oneself as a speaker of Chinese indistinguishable from other native speakers. This is obviously an important empirical question.

5. Our discussion seemed to indicate that a new type of study would be called for -- namely, one in which we could select three or four groups of language students (with perhaps 10 to 20 individuals representing a broad range of proficiency) from each of four language backgrounds -- for illustrative purposes, let us say: (l) French or Spanish; (2) English; (3) Chinese or Japanese; and (4) some other diglossic language. The notion would be to collect at least three types of information from each student -- namely proficiency information using the ILR/ETS/ACTFL assessment procedures preferably for all four skill areas; some other type of multiple indicators of assessment -- using an approach or approaches fundamentally different from ILR/ETS/ACTFL approach to theoretical construct of proficiency and measurement; and a good deal of personal background information concerning the individual and previous language learning/ language use experiences. This research should be conducted by an interdisciplinary team of researchers including members trained in ethnography, second language acquisition

research, language testing, psychometrics, and other areas. Participants might be clients from federal agencies, from academia, or even from the private sector. This research agenda would be a collaborative one with ground rules agreed upon beforehand implemented carefully, and analyzed jointly.

6. Surprisingly, attention was not focused on an increasingly important topic -- language skill attrition or loss. I believe it would be important to examine the question of how "robust" language skill development would be for students at varying levels of proficiency who undergo varying periods of disuse. For example, what is the effect of non-use of the language over a summer break for a student at the 1 level, at the 3 level, etc.? This question has both theoretical as well as practical implications.

7. There exists, of course, a great deal of need for additional psycholinguistic research -- particularly research into the investigation of the acquisition/learning of non-Western languages both as mother tongues and as second languages. It should be noted, however, that a good deal of such research does go on outside North America which is not well known to North American researchers (cf., the issue of *Studies in Second Language Acquisition* devoted to second language acquisition research conducted by French scholars).

8. There exists a need to examine alternative methods of testing within the ILR/ETS/ACTFL framework. John Clark and others have, for example, done exploratory work with a semi-direct testing procedure. Additional work should be done to probe the practicality, reliability, and validity of such a technique. Additionally, one might want to examine in a very systematic fashion the possibility and consequences of oral proficiency assessment administered via telephone, and via videophone.

9. The last exemplary research project which I propose would involve collaboration among an interdisciplinary team of linguists, social psychologists, and test development specialists. What I have in mind is a project analogous to that conducted at CAL by Walt Wolfram, Donna Christian and Debbie Hatfield (1986) involving the description of Vietnamese English as spoken in a community of adolescent and young adult newcomers to the United States. Basically, the have described the non-mainstream variety of English acquired by individuals residing in a heavily "impacted" area of suburban Washington. They describe a pattern of language development and acquisition which is systematic, but yet different from the mainstream variety . There are many regions in which linguists have worked carefully to describe the development of one or another "interlanguages" but to my knowledge, rarely if ever have the subjects who have served as their informants been systematically assessed using the ILR/ETS/ACTFL procedures. I suggest that it would be useful to do so, and that it would also be useful to use tape-recorded samples of the speech to elicit listener judgments in a variation of the matched-guise procedure referred to earlier. One would then have a reasonably precise description of the "guise" which the informant was using, stereotypic reactions by selected groups of listeners concerning the occupational, educational, social or other personal consequences of presenting yourself as a speaker of that guise, and an assessment using the ILR/ETS/ACTFL guidelines which taken together might make a very interesting package for analysis.

These thoughts serve, of course, only to identify exemplary areas for future research. My major concern is that serious thought be given to developing a long-term, proactive research agenda which of necessity will involve collaboration among government personnel and academia by individuals trained in various disciplinary backgrounds. Such an activity could well form a major "agenda" piece for the newly-founded National Foreign Language Center at Johns Hopkins University. It might also be the case that ACTFL or ETS who administer such tests and train testers, could devote a small-but-stable percentage of the income from such testing activities to a pool to be drawn upon to support research. Although that pool of money in and of itself initially would probably not be large enough to support the full cost of research activities, it could very well serve as an example of a commitment to cost-sharing that would encourage government funding agencies or private philanthropic foundations to contribute.

During the course of our discussions a few "special problems" were identified about which mention should be made, but which don't seem to fit neatly into any particular category. A brief listing follows:

1. The fact that the time required for the "typical" French versus the typical Chinese teacher to get a student to the novice-mid level may be widely different is perceived as a problem by some. Is this a problem inherent to the scale or to the setting of the cut point? Is it problematic at all?

2. Does norm shifting occur in use of the rating scale by raters working from 0 to 2+ as opposed to those working in the 3 to 5 range?

3. Is the educated native speaker the appropriate normative reference point? What should the norm be for Arabic -- modern standard Arabic or a colloquial form indigenous to the country in which the student will go to work or to study? It was asserted that we all strive to acquire multiple norms. What norm should we use as a baseline for our assessment? (Here there was a good deal of healthy, but nonconclusive, discussion concerning diglossic situations, the role of code switching in countries such as India, the necessity to master an excessive amount of ritualistic or formulaic language in many of the African language situations, etc.

4. We noted that, in our increasingly litigious society, we must be certain to obtain user judgments concerning the various cut points which are established for decision making. We raised but did not answer the question of how much uncertainty can be afforded and where that uncertainty should be.

5. We talked about the utility of an assessment instrument in providing pedagogical feedback to student or to teacher. There was a reasonably sharp debate concerning whether a test instrument such as the ILR/ETS/ACTFL could in fact be "diagnostic."

6. One topic to which we did not return was that raised by Dan Douglas concerning the importance of emerging technology in the evaluation of foreign language proficiency -- specifically computerized-adaptive testing and the use of interactive videodisc. These techniques for delivery of language assessment procedures will become increasingly prominent and deserve special study in their own right.

Information dissemination

Lastly, let me emphasize the need to actively engage in a systematic program of information dissemination concerning diverse aspects of language proficiency assessment. We need to share information as broadly as possible with colleagues, with clients, and with interested others about techniques for assessing foreign language proficiency. Individuals are interested in at least three types of questions; how well do we as a society do in preparing our students; how well can we do when we enroll such students in special or targeted programs; what types of assessment devices or techniques are available and what are their respective strengths and weaknesses? We have heard that the so-called guidelines are in the process of continual revision. Needless to say, this points to the need for the broadest possible program of information dissemination.

Likewise, the last two presentations of the Symposium -- those which described experiences at the University of Pennsylvania and the University of Minnesota -- need to be documented and distributed. We were fortunate to have representatives from two exemplary programs as participants in this symposium; but as one of the resource people pointed out there are numerous innovative programs in diverse settings throughout North America. There exists a need to systematically identify, describe, and distribute information about such exemplary innovative programs throughout the country. In so doing we should ask what type assessment has accompanied the claim of program success and we should examine whether the measurement techniques used have broader applicability. Finally, we should remember that we must continue to share information systematically with colleagues from around the world. Their needs, their programs, and their approaches are similar to ours and, in some settings, considerable progress has been made which should be systematically exchanged.

Reference

Wolfram, W., Christian, D., & Hatfield, D. (1986). The English of adolescent and young adult Vietnamese refugees in the United States. *World Englishes*, *5*, 47-60.